Problems in Prediction

Andrew James Rieck

November, 2001

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University.

Declaration

This thesis emanates from Andrew Rieck's research only, with all other material explicitly referenced. Andrew Rieck declares that the results contained in this thesis have not been submitted for any other degree.

Andrew James Rieck

Acknowledgements

As for any project, I am grateful to those people who provided moral and technical support during this undertaking. Special and deeply personal thanks go to J. T., A. J., A. P., C. N., P. F., M. M., P. M., and my supervisor, Professor Peter G. Hall.

The former Graduate Convenor of Mathematical Sciences, John Cossey, deserves mention for counseling me through a vicissitude of life immediately prior to undertaking this project. Thanks go to the staff and students at the School of Mathematical Sciences. Lastly, thanks to my family for their love and support.

Abstract

The role of predictive inference is to provide indicative information about a predictand. Typically, this takes the form of a predictor or a prediction region. While the former provides an unambiguous point specification for the predictand, the latter can be perceived to be more informative since its construction is conjoined with prior probability assignment for the predictand.

This thesis defines and reviews the construction and properties of prediction regions. Parametric and nonparametric populations are considered. Methods used to construct prediction regions with identical canonical form are predominantly assessed via coverage error properties.

For a nonparametric population, methods used to construct prediction intervals are suggested for reducing coverage error. The jackknife and smoothed bootstrap are investigated for calibration of prediction intervals; the jackknife performs poorly in that it increases coverage error by an order of magnitude, while the smoothed bootstrap is successful at further reducing coverage error.

A simulation study is conducted to investigate the small sample properties of the former prediction intervals.

Contents

List of Figures

Chapter 1

Introduction

The probability distribution function is a fundamental concept in statistics. If X is an m-dimensional random vector on a probability space $(\Omega, \mathcal{F}, P)$, its probability distribution function F (Chow and Teicher, 1997, Section 6.3) allows probabilities associated with X to be found from the relationship

$$P(X \in A) = \int \mathbf{I}_A(x)\, dF(x)$$

for any $A \in \mathcal{B}^m$, the class of m-dimensional Borel sets, where $x \in \mathbb{R}^m$ and

$$\mathbf{I}_A(x) = \begin{cases} 0 & \text{for } x \notin A, \\ 1 & \text{for } x \in A. \end{cases}$$

Therefore knowledge of the function F allows the construction of a region, denoted by $\mathcal{R}_\alpha$, such that $P(X \in \mathcal{R}_\alpha) = \alpha$ for any $\alpha \in (0, 1)$. When X is a random variable the region $\mathcal{R}_\alpha$ may reduce to an interval $\mathcal{I}_\alpha$ which analogously satisfies $P(X \in \mathcal{I}_\alpha) = \alpha$.

Suppose that the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denotes a sequence of n observed data points which are assumed to be drawn from a population which has the predictand X as an as yet unrealised future observation with an unknown probability distribution function F. The prediction problem, as discussed in this thesis, is the construction of a region or an interval, denoted by $\widehat{\mathcal{R}}_\alpha$ or $\widehat{\mathcal{I}}_\alpha$ respectively, which depend on the sample $\mathcal{X}$, n, and α only, such that

$$P(X \in \widehat{\mathcal{R}}_\alpha) = \alpha \qquad \text{or} \qquad P(X \in \widehat{\mathcal{I}}_\alpha) = \alpha \tag{1.1}$$

for any $\alpha \in (0, 1)$. Then $\widehat{\mathcal{R}}_\alpha$ and $\widehat{\mathcal{I}}_\alpha$ are an exact α-level prediction region and interval, respectively.

Construction of a prediction region or interval can be divided into parametric and nonparametric approaches. For a parametric approach, the predictand X is assumed to have a probability distribution function from one of the known parametric families of distributions, for example the normal distribution with mean μ and variance σ^2. The sample $\mathcal{X}$ is then used to construct a prediction region or interval which does not depend on the unknown parameters of the parametric family. The construction occurs via a pivotal transformation, or a predictive likelihood, function or density. For a nonparametric approach, less rigid assumptions are made about the distribution of the predictand X. Although in the most severe case it will be assumed that various higher order derivatives of F are bounded in an appropriate neighbourhood and that $E(\|X\|^l) < \infty$ for l sufficiently large, the objectivity of the sample $\mathcal{X}$ alone will be used to conduct predictive inference (more than would be the case if F were constrained to lie in a given parametric family of distributions).

Construction of a prediction region or interval proceeds by using the sample $\mathcal{X}$ to estimate population quantities. For a parametric population an exact α-level prediction region or interval can be constructed whenever an appropriate pivotal transformation exists. For a nonparametric population an exact one-sided α-level prediction interval can be constructed whenever the level $\alpha = i/(n+1)$, where the integer $1 \leq i \leq n$, under appropriate population assumptions. To investigate the properties of a proposed prediction region or interval an amenable alternative to (1.1) is required.

For a proposed prediction region and interval, $\widehat{\mathcal{R}}_\alpha$ and $\widehat{\mathcal{I}}_\alpha$ respectively, which depends on the sample $\mathcal{X}$, n, and α only, one natural extension of (1.1) is obtained by allowing

$$P(X \in \widehat{\mathcal{R}}_\alpha) = \alpha + a_n \qquad \text{and} \qquad P(X \in \widehat{\mathcal{I}}_\alpha) = \alpha + b_n,$$

where a_n and b_n are the coverage errors of $\widehat{\mathcal{R}}_\alpha$ and $\widehat{\mathcal{I}}_\alpha$, respectively, and typically a_n, $b_n \to 0$ as $n \to \infty$. The orders of a_n and b_n then represent bounds on the rate, in an asymptotic sense, at which the nominal level α is attained when the amount of information, in this case the sample size n, is large. The nominal α-level prediction region and interval $\widehat{\mathcal{R}}_\alpha$ and $\widehat{\mathcal{I}}_\alpha$, respectively, may then be liberal or conservative for fixed n.

Requirement (1.1) alone will not uniquely determine an exact α-level prediction region even when its construction is possible. As for a proposed nominal α-level prediction region, an exact α-level prediction region can be rendered unique by specifying its canonical form. Two uniquely

identifying canonical forms are, for example, a one-sided prediction interval (assuming the predictand is a random variable), or a prediction region obtained by profiling a function which depends on the sample $\mathcal{X}$ and α. The type of canonical form determines how the probability mass α should be shared, in an asymptotic sense for a nominal α-level prediction region, throughout the probability space.

An outline of this thesis is as follows:

Chapter 2 defines and reviews the methods used to construct a nominal α-level prediction region or interval for a parametric population. One method uses pivotal transformations to construct an exact α-level prediction region for a location-scale population. Other methods rely on the definition of a predictive likelihood, function or density which are analogous, in certain regards, to the conditional probability density function of the predictand given the sample, except that they depend on the sample and α only and not on any unknown population parameters.

Chapter 3 defines and reviews the methods used to construct a nominal α-level prediction interval for a nonparametric population. Firstly, methods based on an independent and identically distributed sample are discussed. These include a Studentised method and quantile estimation methods for a predictand and the percentile-t and accelerated bias-correction methods for a predictand statistic. Secondly, methods are proposed for regression and structural models.

Chapter 4 defines quantile estimation methods used to construct nominal α-level prediction intervals for a nonparametric population. The quantile estimates are constructed from interpolation among quantiles of the empirical distribution function (or equivalently, interpolation among order statistics). Two forms of predictive interval calibration, the jackknife and smoothed bootstrap, are investigated.

Chapter 5 investigates, for small sample sizes, the numerical properties of coverage error for prediction intervals defined in Chapter 4. The convergence of numerical approximations is also considered.

Notes on Notation. The transpose of a vector v and a matrix M are denoted by $v^\top$ and $M^\top$, respectively. The trace, determinant, and inverse of a matrix M are denoted by $\mathrm{tr}M$, $\det M$, and M^{-1}, respectively.

The complement of the set S is denoted by S^c. For example, if $\mathcal{E} \subseteq \Omega$ then

$$\mathcal{E}^c = \{\omega \in \Omega : \omega \notin \mathcal{E}\}.$$

Let $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$ denote a real and a positive real sequence, respectively, of constants. Then $a_n = o(1)$ if $a_n \to 0$ as $n \to \infty$, and $a_n = O(1)$ if $|a_n| < C$ whenever $n \geq n_0$ for some positive real constant C and a fixed integer n_0 greater than one. Additionally, $a_n = o(b_n)$ if and only if $a_n/b_n = o(1)$, and $a_n = O(b_n)$ (a_n is of order b_n) if and only if $a_n/b_n = O(1)$.

Let $\{\Delta_n : n \geq 1\}$ denote a sequence of random variables. Then $\Delta_n = o_p(1)$ (Δ_n converges in probability to zero) if for every $\epsilon > 0$

$$P(|\Delta_n| \geq \epsilon) = o(1),$$

and $\Delta_n = O_p(1)$ if for every $\epsilon > 0$ there exists $\delta(\epsilon) \in (0, \infty)$ such that

$$P\{|\Delta_n| > \delta(\epsilon)\} < \epsilon$$

whenever $n \geq n_0$. Additionally, $\Delta_n = o_p(b_n)$ if and only if $\Delta_n/b_n = o_p(1)$, and $\Delta_n = O_p(b_n)$ (Δ_n is of order b_n in probability) if and only if $\Delta_n/b_n = O_p(1)$.

If Q denotes a population quantity, then $\widehat{Q}$ denotes the version of Q in which the population quantity has been replaced by an estimator determined by the sample $\mathcal{X}$. Following this prescription, a generic theoretical (in that it depends on population quantities) α-level probability region or interval is denoted by $\mathcal{R}_\alpha$ and $\mathcal{I}_\alpha$, respectively, with the corresponding generic nominal (in that population quantities have been replaced by sample quantities) α-level prediction region or interval denoted by $\widehat{\mathcal{R}}_\alpha$ and $\widehat{\mathcal{I}}_\alpha$, respectively.

Note also that no distinction is made between a probability mass function and a probability density function; the latter is used throughout.

In Chapter 2 index notation is employed to simplify expressions (Barndorff-Nielsen and Cox, 1989, Section 5.3). Two conventions are as follows. Firstly, if an index occurs more than once in a term, summation over that index will be assumed. Secondly, $[m]$ after a symbol will indicate a sum of m similar terms, determined by suitable permutations of the indices. For example, if the

indices i, j, and k run from 1 to n then,

$$a_{ik}b_{kji} = \sum_{i,k=1}^{n} a_{ik}b_{kji},$$

$$a_i b^i = \sum_{i=1}^{n} a_i b^i,$$

$$a_{i,jk}[3] = a_{i,jk} + a_{j,ik} + a_{k,ji},$$

and

$$a_{i,jk}[3] + b_{ijk} = \sum_{i,j,k=1}^{n} (a_{i,jk} + a_{j,ik} + a_{k,ji} + b_{ijk}).$$

Additionally, for given m and with $R = r_1 r_2 \cdots r_n$ denoting an arbitrary index set,

$$\sum_{R/m} H(R_1, R_2, \ldots, R_m)$$

will indicate the sum of $H(R_1, R_2, \ldots, R_m)$ over all partitions of R into m blocks with order carrying no significance.

Following Magnus and Neudecker (1999, pp. 87, 100), if $f(x, y) \in \mathbb{R}$ with $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^k$, then $\nabla_y f(x, y)$ and $H_y f(x, y)$ denote the $k \times 1$ gradient vector and the $k \times k$ Hessian matrix of $f(x, y)$ with respect to y.

Chapter 2

Parametric Approaches

In this chapter, parametric approaches to the problem of constructing a prediction region are defined and reviewed. In a Bayesian framework this construction is via the conditional posterior predictive density; in a Frequentist framework this construction is via either a pivotal transformation or a predictive likelihood, function or density. If in a Bayesian framework the correct prior probability distribution is assumed, it is always possible to construct an exact α-level prediction region; an exact α-level prediction region is available in a Frequentist framework whenever a pivotal transformation exists, while a nominal α-level prediction region is derived when a predictive likelihood, function or density is employed.

Let the $n \times m$ sample matrix $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote a sequence of n random vectors which are of length m and in general dependent. Denote a realisation of the sample matrix $\mathcal{X}$ by $\mathfrak{x} = (x_1, x_2, \ldots, x_n)^\top$, where x_i denotes a realisation of the random vector X_i for $i = 1, 2, \ldots, n$, and the probability density function of $\mathcal{X}$ by $f_{\mathcal{X}}(\mathfrak{x}\,;\theta)$, where the parameter vector $\theta \in \Theta$ is of length k and Θ is the parameter space. Let the predictand X denote an as yet unrealised future random vector, not necessarily of the same length as X_i, which is in general dependent on the sample matrix $\mathcal{X}$. Denote a realisation of X by x and assume that the joint probability density function of the sample matrix $\mathcal{X}$ and the predictand X is given by $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$ and that the probability density function of X is given by $f_X(x\,;\theta)$. Additionally, let $f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x}) = f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)/f_{\mathcal{X}}(\mathfrak{x}\,;\theta)\mathbf{I}_{\mathcal{S}}(\mathfrak{x})$ denote the conditional probability density function of X given $\mathcal{X} = \mathfrak{x}$, where $\mathcal{S}$ denotes the support of $f_{\mathcal{X}}(\mathfrak{x}\,;\theta)$.

Suppose that the parameter vector θ has true value θ_0. Knowledge of the true population

probability density function $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta_0)$ allows the construction of an exact α-level prediction region $\widehat{\mathcal{R}}_\alpha$, which is derived via the sample matrix $\mathcal{X}$ and $\alpha \in (0,1)$ only, and satisfies

$$P(X \in \widehat{\mathcal{R}}_\alpha) = \iint \mathbf{I}_{\widehat{\mathcal{R}}_\alpha}(x) f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta_0)\, d\mathfrak{x}\, dx = \alpha. \tag{2.1}$$

When X is a random variable and the prediction region $\widehat{\mathcal{R}}_\alpha$ reduces to the one-sided prediction interval

$$\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{q}_\alpha], \tag{2.2}$$

where it is assumed that $\hat{q}_\alpha \in \mathbb{R}$ depends on the sample matrix $\mathcal{X}$ and α only, requirement (2.1) is sufficient to uniquely define $\widehat{\mathcal{I}}_\alpha$. In this case, $\hat{q}_\alpha$ is given by the conditional α-th quantile of X given $\mathcal{X} = \mathfrak{x}$ and satisfies $\int_{-\infty}^{\hat{q}_\alpha} f_{X|\mathcal{X}}(x\,;\theta_0 \,|\, \mathcal{X})\, dx = \alpha$. In general, further specification of how the probability mass α is shared throughout the probability space is required for $\widehat{\mathcal{R}}_\alpha$ to be uniquely defined. To illustrate this claim by example, let

$$\widehat{\mathcal{R}}_\alpha = \{x : f_{X|\mathcal{X}}(x\,;\theta_0 \,|\, \mathcal{X}) > c_\alpha\} \tag{2.3}$$

be defined by profiling the conditional distribution of the predictand X given the sample matrix $\mathcal{X}$. The real number c_α should be selected such that $\widehat{\mathcal{R}}_\alpha$ satisfies (2.1). Then $\widehat{\mathcal{R}}_\alpha$ is uniquely defined by selecting c_α such that

$$\int_{\widehat{\mathcal{R}}_\alpha} f_{X|\mathcal{X}}(x\,;\theta_0 \,|\, \mathcal{X})\, dx = \alpha.$$

In general, the prediction interval $\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{q}_\alpha]$ and the prediction region $\widehat{\mathcal{R}}_\alpha$ will depend on the true parameter vector θ_0. Methods used to construct an α-level prediction region $\widehat{\mathcal{R}}_\alpha$ in a Bayesian and Frequentist framework when the parameter vector θ is unknown, proceed as follows.

Bayesian Framework. In a fully Bayesian framework, the unknown parameter vector θ is assumed to be a random vector which is coupled with a prior probability density function. This assumption allows the direct evaluation of the conditional posterior predictive density which does not depend on the parameter vector θ. Various methods have been proposed which employ the conditional posterior predictive density to construct an appropriate prediction region. In the following discussion, attention will be focused on the construction of a prediction region and its coverage error properties alone. For a review of various Bayesian approaches used for the construction of

prediction regions, which are not pertinent to this thesis and rely on the specification of a utility function or on general notions of coverage, the interested reader is referred to Aitchison and Dunsmore (1975), Bolfarine and Zacks (1992), and Geisser (1993).

Since the conditional posterior predictive density does not depend on the unknown parameter vector θ, the construction of a prediction region in a fully Bayesian framework proceeds in an analogous way to the case when the unknown parameter vector θ is known and given by θ_0. Details associated with the referred to construction of a prediction region are as follows.

Let the random parameter vector θ be coupled with the prior probability density function $f(\theta)$. Using a form of Bayes' Theorem the conditional posterior predictive density of X given $\mathcal{X} = \mathfrak{x}$, $f(x \mid \mathfrak{x})$, satisfies

$$f(x \mid \mathfrak{x}) = \int f(x \mid \mathfrak{x}, \theta) f(\theta \mid \mathfrak{x}) \, d\theta, \tag{2.4}$$

where

$$f(\theta \mid \mathfrak{x}) = \frac{f(\mathfrak{x} \mid \theta) f(\theta)}{f(\mathfrak{x})}$$

and

$$f(x \mid \mathfrak{x}, \theta) = f_{X \mid \mathcal{X}}(x; \theta \mid \mathfrak{x}), \; f(\mathfrak{x} \mid \theta) = f_{\mathcal{X}}(\mathfrak{x}; \theta), \; f(\mathfrak{x}) = \int f(\mathfrak{x} \mid \theta) f(\theta) \, d\theta. \tag{2.5}$$

The conditional posterior predictive density can be used to construct an α-level prediction region. Two methods are as follows.

Set $\widehat{\mathcal{R}}_\alpha = \{x : f(x \mid \mathcal{X}) > c_\alpha\}$ where the nonnegative real constant c_α is selected such that $\int_{\widehat{\mathcal{R}}_\alpha} f(x \mid \mathcal{X}) \, dx = \alpha$. When X is a random variable, set $\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{q}_\alpha]$, where $\hat{q}_\alpha$ denotes the α-th quantile of the conditional posterior predictive density of X given $\mathcal{X} = \mathfrak{x}$ and satisfies $\int_{-\infty}^{\hat{q}_\alpha} f(x \mid \mathcal{X}) \, dx = \alpha$. Then, setting $f(x, \mathfrak{x} \mid \theta) = f_{(X,\mathcal{X})}(x, \mathfrak{x}; \theta)$ and $f(x, \mathfrak{x}) = \int f(x, \mathfrak{x} \mid \theta) f(\theta) \, d\theta$, it follows that

$$\begin{aligned} P(X \in \widehat{\mathcal{R}}_\alpha) &= \iint \mathbf{I}_{\widehat{\mathcal{R}}_\alpha}(x) f(x, \mathfrak{x}) \, dx \, d\mathfrak{x} \\ &= \iint \mathbf{I}_{\widehat{\mathcal{R}}_\alpha}(x) f(x \mid \mathfrak{x}) f(\mathfrak{x}) \, dx \, d\mathfrak{x} \\ &= \alpha \int f(\mathfrak{x}) \, d\mathfrak{x} = \alpha. \end{aligned}$$

Similarly, $P(X \in \widehat{\mathcal{I}}_\alpha) = \alpha$. Therefore, in a Bayesian framework, $\widehat{\mathcal{R}}_\alpha$ is an exact α-level prediction region and $\widehat{\mathcal{I}}_\alpha$ is an exact α-level prediction interval.

While the prior probability density function for the parameter vector θ allows the calculation of an exact α-level prediction region, it also goads the statistician towards alternative approaches which do not require a prior probability density function to be specified. While in some instances logical considerations may guide selection of the prior probability density function, there is, in general, no agreement on specification of a prior probability density function. It will be seen below that the conditional posterior predictive density influences the methods which are proposed in a Frequentist framework where it is assumed the parameter vector θ is an unknown constant vector that can be estimated via the sample matrix $\mathcal{X}$.

Frequentist Framework. The remainder of this chapter will concentrate on various methods used to construct an α-level prediction region in a Frequentist framework. The discussion is organised as follows.

Section 2.1 illustrates a pivotal transformation from which an exact α-level prediction region can be constructed whenever the sample $\mathcal{X}$ and the predictand X belong to a location-scale population.

An alternative method is to define a function of the sample $\mathcal{X}$ and the predictand X which is used in an analogous way to the conditional distribution of X given $\mathcal{X} = \mathfrak{x}$, assuming the parameter vector θ has known true value θ_0, to construct a nominal α-level prediction region (see (2.2) or (2.3)). Candidates for the former function are reviewed as follows.

Section 2.2 introduces various predictive likelihoods. The sufficiency based predictive likelihoods obviate the presence of the parameter vector θ by conditioning with respect to a sufficient statistic. In contrast, the approximate predictive likelihood is derived from an asymptotic expansion of the conditional posterior predictive density, where it is assumed that the prior density of the parameter vector θ is constant on its support. Alternatively, the approximate conditional predictive likelihood and the modified profile predictive likelihood are based on the joint probability density function of $\mathcal{X}$ and X, $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$, where the parameter vector θ is replaced by an estimator based on the sample $\mathcal{X}$ and the predictand X, thereby following an indirectly proposed technique of Berger and Wolpert (1984).

Section 2.3 constructs predictive functions which are determined from an axiomatic foundation.

Section 2.4 proposes two types of predictive densities. The first type is obtained by investigating estimators of the true conditional distribution of X given the sample $\mathcal{X} = \mathfrak{x}$, $f_{X|\mathcal{X}}(x\,;\theta_0\,|\,\mathfrak{x})$; the second type is retrospectively determined from asymptotic considerations of the upper end-point

of the exact one-sided α-level prediction interval that is based on the sample $\mathcal{X}$ only.

Section 2.5 investigates the explicit coverage error calculation for derived prediction regions of various methods. Analytic and bootstrap calibration of a prediction region are also considered; a predictive density is retrospectively determined from the analytic calibration of a one-sided prediction interval and is a function of the sample $\mathcal{X}$ and α only.

Properties of a Predictive Likelihood, Function or Density. Consider now the properties which may be intrinsic to a predictive likelihood, function or density. Let $p(x\,;\mathfrak{x})$ denote any function of the predictand X and the sample $\mathcal{X}$ which may be used to construct predictive regions. The relative conceptual advantages of $p(x\,;\mathfrak{x})$ over various competitors can be assessed via invariance properties in conjunction with the computational convenience and coverage error properties of its derived prediction region. Invariance properties categorise the derived similarities in $p(x\,;\mathfrak{x})$ resulting from changes to the parameter vector θ, the sample $\mathcal{X}$ or the predictand X. Three such changes occur when θ, $\mathcal{X}$, and X are replaced by $h_1(\theta)$, $h_2(\mathcal{X})$, and $h_3(X)$, respectively, where h_1, h_2, and h_3 are smooth one-to-one transformations. Should $p(x\,;\mathfrak{x})$ remain unchanged if the parameter vector θ is replaced by $h_1(\theta)$ in the joint probability density function $f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\cdot)$, the sample $\mathcal{X}$ is replaced by $h_2(\mathcal{X})$ in $p(x\,;\cdot)$, or the predictand X is replaced by $h_3(X)$ in $p(\cdot\,;\mathfrak{x})$, then $p(x\,;\mathfrak{x})$ is said to be parameter, sample, or predictand invariant, respectively. Additionally, $p(x\,;\mathfrak{x})$ is referred to as being scale invariant when it is predictand invariant with $h_3(X) = cX$ for some real constant $c \neq 0$.

2.1 Pivotal Transformations

This section illustrates how a pivotal transformation can be utilised to construct an exact α-level prediction region for a parametric population; a detailed account of this method will be given for a location-scale population. A condition which explicitly determines a property of a prediction region is also discussed.

Let $g(\mathcal{X},\, X)$ denote any function of the sample $\mathcal{X}$ and the predictand X, where $\mathcal{X}$ and X have a joint probability density function given by $f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)$ with unknown parameter vector $\theta \in \Theta$. Then g is said to be a pivotal transformation for the predictand X if the distribution of $g(\mathcal{X},\, X)$, evaluated under the distribution of $\mathcal{X}$ and X when the parameter vector is θ, is the same

for all $\theta \in \Theta$. In this way, g does not depend on the parameter vector θ. The notion of a pivotal transformation was first introduced by R. A. Fisher (Fisher, 1934; Fisher, 1935); note that Fisher's definition differs from contemporary usage (Barnard, 1985).

When the predictand X is a random variable, Barndorff-Nielsen and Cox (1994) stress the importance of the pivotal transformation $g(\mathcal{X}, X)$ being monotone in X for almost all $\mathcal{X}$. While monotonicity is not an essential requirement, it does allow the convenient construction of a prediction region through an appropriate quantile of $g(\mathcal{X}, X)$ in conjunction with its inversion with respect to X. When the pivotal transformation is not monotone in X, more specialised techniques of inversion may be required for prediction region construction (Reiss, 1989, Appendix 1).

To demonstrate a pivotal transformation, let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables from a normal population which has probability density function given by $f(x) = \exp\{\frac{1}{2}(x-\mu)^2/\sigma_0^2\}/(\sigma_0\sqrt{2\pi})$ with unknown mean μ and known variance $\sigma_0^2 > 0$. The predictand X is independent of the sample $\mathcal{X}$ but drawn from the same population. Set $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and

$$g(\mathcal{X}, X) = \frac{X - \bar{X}}{\sigma_0\sqrt{1+\frac{1}{n}}}.$$

Then g is a pivotal transformation since it has a standard normal distribution for all $\mu \in \mathbb{R}$.

Let z_α denote the α-th quantile of a standard normal distribution function. Then, since g is monotone in X, it follows that

$$\widehat{\mathcal{I}}_\alpha = \left(-\infty, \bar{X} + z_\alpha\sigma_0\sqrt{1+n^{-1}}\right]$$

is an exact one-sided α-level prediction interval for the predictand X since

$$\begin{aligned} P(X \in \widehat{\mathcal{I}}_\alpha) &= P\left\{X \le \bar{X} + z_\alpha\sigma_0\sqrt{1+n^{-1}}\right\} \\ &= P\left\{\frac{X-\bar{X}}{\sigma_0\sqrt{1+n^{-1}}} \le z_\alpha\right\} = \alpha. \end{aligned}$$

If in addition the normal population also has an unknown variance $\sigma^2 > 0$, set

$$\hat{S}^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

and

$$g(\mathcal{X}, X) = \frac{X - \bar{X}}{\hat{S}\sqrt{1+\frac{1}{n}}}. \tag{2.6}$$

Then g is a pivotal transformation since it has Student's t distribution with $n-1$ degrees of freedom for all $\theta = (\mu, \sigma^2)^\top$.

Let $t_{\alpha,n-1}$ denote the α-th quantile of Student's t distribution function with $n-1$ degrees of freedom. Then, since the function g is monotone in X, it follows that

$$\widehat{\mathcal{I}}_\alpha = \Big(-\infty, \bar{X} + t_{\alpha,n-1}\hat{S}\sqrt{1+n^{-1}}\Big],$$

is an exact one-sided α-level prediction interval for the predictand X.

A more general parametric class than the set of all normal populations is the set of all location-scale populations of a given type. If W is a random variable from a location-scale population it has a probability density function

$$f_W(w\,;\mu,\sigma) = \sigma^{-1}h\big\{(w-\mu)/\sigma\big\},$$

where both μ and $\sigma > 0$ are real unknowns, w denotes a realisation of the random variable W, and h is a known probability density function.

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables drawn from a location-scale population. Again the predictand X is independent of $\mathcal{X}$ but drawn from the same population. Let $\mathbf{1}$ denote the column vector of length n (the sample size of $\mathcal{X}$) whose components all equal one. Then the joint probability density function of $\mathcal{X}$ and X is given by

$$f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\mu,\sigma) = \sigma^{-n-1}h\big\{(x-\mu)/\sigma\big\}\Pi_{i=1}^n h\big\{(x_i-\mu)/\sigma\big\}.$$

Observe that $Y = \sigma^{-1}(\mathcal{X} - \mu\mathbf{1})$ and $Z = \sigma^{-1}(X-\mu)$ have a joint probability density function (not depending on μ and σ) given by

$$f_{(Y,Z)}(y,z) = h(z)\Pi_{i=1}^n h(y_i),$$

where $y = (y_1, y_2, \ldots, y_n)^\top$ and z denotes a realisation of the random vector Y and the random variable Z, respectively.

Consider the random vector transformation of $(Y^\top, Z)^\top$ to $(R, S, T, \breve{U}^\top)^\top$, with

$$Y = S(T\mathbf{1} + U), \quad Z = S(T+R) \tag{2.7}$$

subject to

$$S \geq 0, \quad U^\top\mathbf{1} = 0, \quad U^\top U = n(n-1), \tag{2.8}$$

where R, S, and T are random variables, $U = (U_1, U_2, \ldots, U_n)^\top$ is a random vector, and $\check{U} = (U_1, U_2, \ldots, U_{n-2})^\top$. Note that two components of U are determined by the restrictions specified at (2.8). Furthermore, the transformation specified by (2.7) and (2.8) is equivalent to

$$R = \frac{n^{1/2}(X - \bar{X})}{\hat{S}}, \; S = \frac{\hat{S}}{n^{1/2}\sigma}, \; T = \frac{n^{1/2}(\bar{X} - \mu)}{\hat{S}}, \; U = \frac{n^{1/2}}{\hat{S}}(\mathcal{X} - \bar{X}). \tag{2.9}$$

The joint density of R, S, T, and $\check{U}$ is given by

$$f_{(R,S,T,\check{U})}(r, s, t, \check{u}) = \frac{n(n-1)s^n}{|u_n - u_{n-1}|} f_{(Y,Z)}\{s(t\mathbf{1} + u), s(t + r)\},$$

where $u = (u_1, u_2, \ldots, u_n)^\top$ and $\check{u} = (u_1, u_2, \ldots, u_{n-2})^\top$ denotes a realisation of the random vectors U and $\check{U}$, respectively. The conditional probability density function of R, S, and T given $U = u$ is

$$f_{(R,S,T)|U}(r, s, t; \mu, \sigma \mid u) = \frac{s^n f_{(Y,Z)}\{s(t\mathbf{1} + u), s(t + r)\}}{\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} s^n f_{(Y,Z)}\{s(t\mathbf{1} + u), s(t + r)\}\, dr\, ds\, dt}.$$

And the marginal distribution of R given $U = u$ is obtained by integrating over s and t, and given by

$$f_{R|U}(r \mid u) = \frac{\int_{-\infty}^{\infty}\int_{0}^{\infty} s^n f_{(Y,Z)}\{s(t\mathbf{1} + u), s(t + r)\}\, ds\, dt}{\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} s^n f_{(Y,Z)}\{s(t\mathbf{1} + u), s(t + r)\}\, dr\, ds\, dt}$$

which does not depend on μ and σ; therefore, conditional on $U = u$, R is a pivotal transformation for the predictand X.

Let $F_{R|U}(r \mid u) = \int_{-\infty}^{r} f_{R|U}(w \mid u)\, dw$ denote the conditional distribution function of R given $U = u$ and denote by v_α the α-th quantile of $F_{R|U}(r \mid u)$. Since R is monotone in X (see (2.9)) it follows that

$$\widehat{\mathcal{I}}_\alpha = (-\infty, \bar{X} + n^{-1/2}\hat{S}v_\alpha] \tag{2.10}$$

is an exact one-sided α-level prediction interval since

$$P(X \in \widehat{\mathcal{I}}_\alpha) = P(R \leq v_\alpha) = E\{F_{R|U}(v_\alpha \,|U)\} = \alpha. \tag{2.11}$$

The former result is derived from the tower property of conditional expectation. That is, let $(\Omega, \mathcal{F}, P)$ denote the underlying probability space where $\mathcal{G}_1$ and $\mathcal{G}_2$ are σ-algebras with $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$. Then for a generic random variable X on the former probability space, it can be shown that

$$E\{E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\} = E\{E(X \mid \mathcal{G}_1) \mid \mathcal{G}_2\} = E(X \mid \mathcal{G}_1), \tag{2.12}$$

with probability one, see Chung (1974, Chapter 9), Williams (1991), and Chow and Teicher (1997, Chapter 7). To derive (2.11) take $\mathcal{G}_1 = \{\emptyset, \Omega\}$, $\mathcal{G}_2 = \sigma(U)$ at (2.12).

One way to rectify the arbitrary shape of a two-sided prediction interval is via probability centering. That is, the probability of the predictand X not exceeding the lower end-point of $\widehat{\mathcal{I}}_\alpha$ equals the probability of the predictand X exceeding the upper end-point. Following this convention, where X and $\mathcal{X}$ defined as at (2.10), an exact two-sided α-level prediction interval for the predictand X is given by

$$\widehat{\mathcal{I}}_\alpha = [\bar{X} + n^{-1/2}\hat{S}v_{(1-\alpha)/2}, \bar{X} + n^{-1/2}\hat{S}v_{(1+\alpha)/2}]. \tag{2.13}$$

When $h(w) = \frac{1}{\sqrt{2\pi}}\exp(-w^2/2)$ is the standard normal density, (2.13) specialises to

$$\widehat{\mathcal{I}}_\alpha = [\bar{X} + n^{-1/2}\hat{S}t_{(1-\alpha)/2,n-1}, \bar{X} + n^{-1/2}\hat{S}t_{(1+\alpha)/2,n-1}].$$

A generalisation of the notion of probability centering in one-dimension, when a pivotal distribution for the random vector predictand X exists or otherwise, is proposed by Beran (1993). The nominal α-level prediction region for the predictand X is of the form

$$\widehat{\mathcal{R}}_\alpha = \{x : Z(u,x) \leq c_\alpha(u,\mathcal{X}),\ \forall u \in U\}, \tag{2.14}$$

where $Z = \{Z(u,X),\ \forall u \in U\}$ is a random process with index set U which is assumed to be a metric space. For some constant $\beta(\alpha,\theta)$ the critical values $c_\alpha(u,\mathcal{X})$ are selected such that

$$P(X \in \widehat{\mathcal{R}}_\alpha|\mathcal{X}) - \alpha = o_p(1) \tag{2.15}$$

and

$$\sup_{u\in U}\left|P(X \in \widehat{\mathcal{R}}_{\alpha,u}|\mathcal{X}) - \beta(\alpha,\theta)\right| = o_p(1), \tag{2.16}$$

where

$$\widehat{\mathcal{R}}_{\alpha,u} = \{x : Z(u,x) \leq c_\alpha(u,\mathcal{X})\},\quad u \in U.$$

Note that (2.15) compels, via the Mean Convergence Criterion (Chow and Teicher, 1997, Section 4.2),

$$P(X \in \widehat{\mathcal{R}}_\alpha) = \alpha + o(1).$$

Additionally, (2.16) specifies the sense in which $\widehat{\mathcal{R}}_\alpha$ is centred by requiring $P(X \notin \widehat{\mathcal{R}}_{\alpha,u}|\mathcal{X})$ to converge in probability to a limit which is identical for every $u \in U$.

To elaborate, let the $n \times m$ sample matrix $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random vectors from an m-variate normal population with $m \times 1$ mean vector μ and $m \times m$ variance matrix Σ with $m < n$. The predictand X is a random vector which is independent of the sample matrix $\mathcal{X}$ but drawn from the same population. Set $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and $\hat{S}^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top$. Then (Anderson, 1958, Theorem 5.2.2) it follows that

$$(X - \bar{X})^\top (\hat{S}^2)^{-1} (X - \bar{X}) \sim \{(1 + n^{-1})(n-1)m\}/(n-m) F_{m,n-m}$$

is a pivotal transformation for the predictand X, where $F_{m,n-m}$ has an F distribution with m and $n - m$ degrees of freedom. Therefore,

$$\widehat{\mathcal{R}}_\alpha = \{x : (x - \bar{X})^\top (\hat{S}^2)^{-1} (x - \bar{X}) \leq d_{\alpha,m,n}\} \tag{2.17}$$

is an exact α-level prediction region, where $d_{\alpha,m,n} = \{(1 + n^{-1})(n-1)m\}/(n-m) F_{\alpha,m,n-m}$ and $F_{\alpha,m,n-m}$ denotes the α-th quantile of an F probability distribution function with m and $n - m$ degrees of freedom.

Let $U = \{u \in \mathbb{R}^m : \|u\| = 1\}$ denote the $(m-1)$-sphere in $\mathbb{R}^m$ centered at the origin. Analogous to results in Miller (1981, Chapter 2), (2.17) is equivalent to

$$\widehat{\mathcal{R}}_\alpha = \{x : u^\top x \leq u^\top \bar{X} + u^\top \hat{S}^2 u \, d_{\alpha,m,n}^{1/2}, \ \forall u \in U\}.$$

The prediction region $\widehat{\mathcal{R}}_\alpha$ for the predictand X is the intersection of the uncountably many prediction half-spaces

$$\widehat{\mathcal{R}}_{\alpha,u} = \{x : u^\top x \leq u^\top \bar{X} + u^\top S_{\mathcal{X}}^2 u \, d_{\alpha,m,n}^{1/2}\}, \qquad u \in U,$$

with

$$P(X \in \widehat{\mathcal{R}}_\alpha | \mathcal{X}) - \alpha = o_p(1)$$

and

$$\sup_{u \in U} \left| P(X \in \widehat{\mathcal{R}}_{\alpha,u} | \mathcal{X}) - \Phi\left[\{G^{-1}(\alpha)\}^{1/2}\right] \right| = o_p(1),$$

where Φ and G are the standard normal and chi-squared, with m degrees of freedom, probability distribution functions, respectively. Therefore, $\widehat{\mathcal{R}}_\alpha$ is given by (2.14) with $Z(u, X) = u^\top X$, $c_\alpha(u, \mathcal{X}) = u^\top \bar{X} + u^\top S_{\mathcal{X}}^2 u \, d_{\alpha,m,n}^{1/2}$, and $\beta(\alpha, \theta) = \Phi[\{G^{-1}(\alpha)\}^{1/2}]$.

Under appropriate regularity conditions the critical value $c_\alpha(u, \mathcal{X})$ can be selected such that the corresponding prediction region $\widehat{\mathcal{R}}_\alpha$ satisfies both (2.15) and (2.16) (Beran, 1993).

In principle, a pivotal transformation for the predictand X allows the construction of an exact α-level prediction region. However pivotal transformations are not available for a large class of problems (Barndorff-Nielsen, 1980, Examples 6 and 7). To amend this situation, various asymptotic approaches are considered which specify nominal α-level prediction regions that are constructed via a predictive likelihood, function or density as defined in Section 2.2, Section 2.3, and Section 2.4.

2.2 Predictive Likelihood

Parametric inference for the parameter vector θ may be based on the parametric likelihood. This section introduces various versions of predictive likelihood which are conceptually perceived as a form of likelihood for the predictand X. In Subsection 2.2.1 the notion of sufficiency is advantageously used to construct predictive likelihoods which do not depend on the unknown parameter vector θ. Motivation for one predictive likelihood based on sufficiency is obtained from the predictive inference version of the relationship between the parametric likelihood in a Frequentist framework and the conditional probability density of θ given $\mathcal{X} = \mathfrak{x}$ in a Bayesian framework. In Subsection 2.2.2 predictive likelihoods are proposed which do not depend on sufficiency. The approximate predictive likelihood is derived from the asymptotic expansion of the posterior predictive density with constant prior. The approximate conditional predictive likelihood and the modified profile predictive likelihood replace θ in $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$ by an estimator based on the sample $\mathcal{X}$ and the predictand X.

2.2.1 Sufficiency Based Predictive Likelihood

Reiterating, the sample matrix $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ has probability density function $f_{\mathcal{X}}(\mathfrak{x}\,;\theta)$. In a Frequentist framework, parametric inference for the unknown parameter vector θ may be based on

$$L(\theta\,;\mathfrak{x}) = f_{\mathcal{X}}(\mathfrak{x}\,;\theta),$$

the parametric likelihood of θ (Barndorff-Nielsen and Cox, 1994). An exact α-level likelihood-based confidence region for the parameter vector θ takes the form

$$\widehat{\mathcal{R}}_{\mathrm{CL},\alpha} = \{\theta : L(\theta; \mathfrak{x}) > c_{\mathrm{L},\alpha}\}, \tag{2.18}$$

where $c_{\mathrm{L},\alpha}$ is a nonnegative real number selected such that $P(\theta_0 \in \widehat{\mathcal{R}}_{\mathrm{CL},\alpha}) = \alpha$.

In a Bayesian framework, parametric inference for the parameter vector θ is based on

$$f(\theta \mid \mathfrak{x}) = \frac{f(\mathfrak{x} \mid \theta) f(\theta)}{f(\mathfrak{x})},$$

the conditional probability density of θ given $\mathcal{X} = \mathfrak{x}$, where $f(\mathfrak{x} \mid \theta)$ and $f(\mathfrak{x})$ are defined at (2.5), and $f(\theta)$ denotes the prior probability density function of θ. An exact α-level Bayesian-based confidence region for θ takes the form

$$\widehat{\mathcal{R}}_{\mathrm{CB},\alpha} = \{\theta : f(\theta \mid \mathfrak{x}) > c_{\mathrm{B},\alpha}\}, \tag{2.19}$$

where $c_{\mathrm{B},\alpha}$ is a nonnegative real number selected such that $P(\theta_0 \in \widehat{\mathcal{R}}_{\mathrm{CB},\alpha}) = \alpha$.

The Bayesian conditional probability density can be written as

$$f(\theta \mid \mathfrak{x}) = g_1(\mathfrak{x}, \theta) L(\theta; \mathfrak{x}), \tag{2.20}$$

where $g_1(\mathfrak{x}, \theta) = f(\theta)/f(\mathfrak{x})$. Hence the role of the parametric likelihood $L(\theta; \mathfrak{x})$ in the construction of both the likelihood and Bayesian based confidence regions, given by (2.18) and (2.19), respectively, is made explicit. Therefore, following a convention of Hinkley (1979), it could be expected that a predictive likelihood for the predictand X, denoted by $p_{\mathrm{L}}(x; \mathfrak{x})$, should satisfy the factorisation

$$f(x \mid \mathfrak{x}) = g_2(\mathfrak{x}, x) p_{\mathrm{L}}(x; \mathfrak{x}), \tag{2.21}$$

where g_2 is determined by the marginal prior distributions of $\mathcal{X}$ and X, and $f(x \mid \mathfrak{x})$ is the conditional posterior predictive density defined at (2.4). In this way the parametric likelihood $L(\theta; \mathfrak{x})$ and the predictive likelihood $p_{\mathrm{L}}(x; \mathfrak{x})$ exhibit analogous roles in relation to their appropriate conditional Bayesian posterior densities (compare (2.20) and (2.21)). On a historical note, it may have been the widespread acceptance of parametric likelihood as a general method for parametric inference that prompted (2.21) to be seen as a fundamental identity which should be satisfied by a predictive likelihood.

Before introducing various versions of predictive likelihood, recall that a statistic $T = T(\mathcal{X})$ of the sample $\mathcal{X}$ with probability density function $f_{\mathcal{X}}(\mathfrak{x}; \theta)$, of a discrete underlying population,

is said to be *sufficient* for $\mathcal{X}$ if the conditional distribution of $\mathcal{X}$ given $T = t$ does not depend on the unknown parameter vector θ for all t. When the underlying population is continuous, causing the non-uniqueness of the conditional probability density, an alternative definition of sufficiency is required to circumvent technical difficulties. The interested reader is referred to Barndorff-Nielsen (1978), Barndorff-Nielsen and Cox (1994), and Lehmann (1997a, 1997b). Additionally, a sufficient statistic $T = T(\mathcal{X})$ is said to be *minimal sufficient* if for any sufficient statistic S there exists a function h such that $T = h(S)$ with probability one; the existence of minimal sufficient statistics has been shown by Bahadur (1954, 1957). Hinkley (1979, 1980) uses the notion of sufficiency to define a predictive likelihood which does not depend on the unknown parameter vector θ, as follows.

Consider the case where the sample $\mathcal{X}$ and the predictand X are independent. Furthermore, assume the existence of minimal sufficient statistics for $(\mathcal{X}, X)$, $\mathcal{X}$, and X which are denoted by R, S, and T, respectively, where it is assumed that R, S, and T provide an appropriate reduction. Additionally, assuming t is uniquely defined by r and s, Hinkley (1979) defines the predictive likelihood for the predictand X as

$$p_{\mathrm{SLH}}(x\,;\mathfrak{x}) = f_{\mathcal{X}\,|\,S}(\mathfrak{x}\,|\,s) f_{X\,|\,T}(x\,|\,t) f_{S\,|\,R}\{s\,|\,r(s,t)\}, \tag{2.22}$$

where $f_{\mathcal{X}\,|\,S}(\mathfrak{x}\,|\,s)$ denotes the conditional probability density function of $\mathcal{X}$ given $S = s$, $f_{X\,|\,T}(x\,|\,t)$ denotes the conditional probability density function of X given $T = t$, and $f_{S\,|\,R}\{s\,|\,r(s,t)\}$ denotes the conditional probability density function of S given $R = r$. According to Bjørnstad (1990), $p_{\mathrm{SLH}}(x\,;\mathfrak{x})$ may be simplified in a fully discrete setting to give

$$p_{\mathrm{SLI}}(x\,;\mathfrak{x}) = f_{(\mathcal{X},X)\,|\,R}\{\mathfrak{x}, x\,|\,r(s,t)\}, \tag{2.23}$$

where $f_{(\mathcal{X},X)\,|\,R}\{\mathfrak{x}, x\,|\,r(s,t)\}$ denotes the conditional probability density function of $(\mathcal{X}, X)$ given $R = r$. Additionally, when the minimal sufficient statistic T equals the predictand X, the predictive likelihood proposed by Lauritzen (1974), and defined by $p_{\mathrm{SLL}}(x\,;\mathfrak{x}) = f_{X\,|\,R}(x\,|\,r)$, is identical to $p_{\mathrm{SLI}}(x\,;\mathfrak{x})$. However, when the minimal sufficient statistic T does not equal X there exist counter-examples (Bjørnstad, 1990, Example 3) in which it is shown that $p_{\mathrm{SLI}}(x\,;\mathfrak{x})$ and $p_{\mathrm{SLL}}(x\,;\mathfrak{x})$ are in general not identical.

Consider the case where the sample $\mathcal{X}$ and the predictand X are dependent. Furthermore, assume the existence of the minimal sufficient statistic and the sufficient statistic for $(\mathcal{X}, X)$ and

$\mathcal{X}$ denoted by R and S, respectively, where it is assumed that R and S provide an appropriate reduction. Additionally let T be a function of (S, X) such that

1. R is determined by (S, T), and,

2. the minimal sufficient reduction of X is determined by (S, T).

Assume that the function $r(s, t)$ has a unique inverse which equals $t(r, s)$ for each fixed value of s. Then Hinkley (1979) defines the predictive likelihood for the predictand X as

$$p_{\mathrm{SLH}}(x\,;\mathfrak{x}) = f_{\mathcal{X}\,|\,S}(\mathfrak{x}\,|\,s) f_{X\,|\,(S,T)}(x\,|\,s,t) f_{S\,|\,R}\{s\,|\,r(s,t)\}, \tag{2.24}$$

where $f_{X\,|\,(S,T)}(x\,|\,s,t)$ denotes the conditional probability density function of X given $(S,T) = (s,t)$.

While Hinkley (1979) and Lauritzen (1974) develop predictive likelihoods p_{SLH} and p_{SLL}, respectively, by employing sufficiency in a direct sense, Butler (1986) considers a circuitous approach, as follows.

Suppose the sample $\mathcal{X}$ and the predictand X are dependent with probability density function given by $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$, where θ is the unknown parameter vector. Assume the existence of the minimal sufficient statistic for $(\mathcal{X}, X)$ which is denoted by R. Additionally, assuming that $R(\mathcal{X}, X)$ provides a reduction for $(\mathcal{X}, X)$, denote by $u(\mathfrak{x}, x)$ the vector of orthogonal coordinates in the $(\mathfrak{x}, x)$-space which is locally orthogonal to $r(\mathfrak{x}, x)$ for all $\mathfrak{x}$ and x such that the transformation from $(\mathfrak{x}, x)$ to (r, u) is one-to-one. The former transformation allows the joint probability density of $\mathcal{X}$ and X, $f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$, to be decomposed into two components. One will involve the unknown parameter vector θ, while the other will not.

To identify this decomposition, let

$$J = \partial r / \partial(\mathfrak{x}, x) \tag{2.25}$$

and

$$K = \partial u / \partial(\mathfrak{x}, x)$$

denote the Jacobian matrices of $r(\mathfrak{x}, x)$ and $u(\mathfrak{x}, x)$. Since $u(\mathfrak{x}, x)$ is locally orthogonal, $KK^{\top} = \mathbf{I}$ and $KJ^{\top} = \mathbf{0}$, the determinant of the Jacobian for the coordinate change from $(\mathfrak{x}, x)$ to (r, u) is

given by

$$\det\left(\frac{\partial(r,u)}{\partial(\mathfrak{x},x)}\right) = \det(JJ^\top)^{1/2}, \tag{2.26}$$

where $\mathbf{I}$ and $\mathbf{0}$ denote the identity and zero matrix, respectively, of appropriate dimension. This follows from the fact that the determinant of a matrix is unchanged by transposition and that the determinant of a product of matrices equals the product of the determinants (Kreyszig, 1999). Using (2.26) the joint probability density $f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)$ can be expressed as

$$h_1(r,u\,;\theta)h_2(r\,;\theta),$$

where

$$h_1(r,u\,;\theta) = \det(JJ^\top)^{-1/2}\frac{f_{(\mathcal{X},X)}\{\mathfrak{x}(r,u),x(r,u)\,;\theta\}}{f_R(r\,;\theta)}$$

and

$$h_2(r\,;\theta) = f_R(r\,;\theta), \tag{2.27}$$

with $f_R(r\,;\theta)$ denoting the probability density function of R which depends on the unknown parametric vector θ.

Since R is the minimal sufficient statistic for $(\mathcal{X},X)$,

$$h_1(r,u\,;\theta) = \det(JJ^\top)^{-1/2}\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{f_R(r\,;\theta)} = \det(JJ^\top)^{-1/2}f_{(\mathcal{X},X)|R}(\mathfrak{x},x\,|r)$$

is more appropriately denoted by $h_1(r,u)$ because it does not depend on the unknown parameter vector θ. Therefore, Butler (1986) defines the predictive likelihood for the predictand X as

$$p_{\mathrm{SLB}}(x\,;\mathfrak{x}) = \det(JJ^\top)^{-1/2}f_{(\mathcal{X},X)|R}(\mathfrak{x},x\,|r) = \det(JJ^\top)^{-1/2}p_{\mathrm{SLI}}(x\,;\mathfrak{x}), \tag{2.28}$$

where $p_{\mathrm{SLI}}(x\,;\mathfrak{x})$, introduced by Bjørnstad (1990), is defined by (2.23).

Since any one-to-one transformation of a minimal sufficient statistic is minimally sufficient, it follows that the predictive likelihoods p_{SLH} and p_{SLI} depend on the choice of the minimal sufficient statistic R. Alternatively, the predictive likelihood introduced by Butler (1986) and denoted by p_{SLB} (see (2.28)) is invariant with respect to the choice of minimal sufficient statistic R. This fact can be verified by letting $\hat{R}$ denote any one-to-one transformation of the minimal sufficient statistic R, where the Jacobian of $\hat{R}$ is given by $\hat{J} = \partial\hat{r}/\partial(\mathfrak{x},x)$. Using the chain rule (Magnus and

Neudecker, 1999), $\hat{J} = WJ$, where $W = \partial\hat{r}/\partial r$ and J is the Jacobian of R (see (2.25)). Suppose that $\hat{R}$ is used as the minimal sufficient statistic for $(\mathcal{X}, X)$. Then

$$\begin{aligned} p_{\mathrm{SLB}}(x\,;\mathfrak{x}) &= \det(\hat{J}\hat{J}^{\top})^{-1/2}\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{f_{\hat{R}}(\hat{r}\,;\theta)} \\ &= \det(W)^{-1}\det(JJ^{\top})^{-1/2}\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{f_{\hat{R}}(\hat{r}\,;\theta)} \\ &= \det(JJ^{\top})^{-1/2}\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{f_{R}(r\,;\theta)}, \end{aligned}$$

which denotes the predictive likelihood of Butler (1986), given by (2.28), with R used as the minimal sufficient statistic for $(\mathcal{X}, X)$. Furthermore, as stressed in Bjørnstad (1990), the predictive likelihood p_{SLI} (see (2.23)) is parameter and scale invariant. Therefore, since (2.28) specifies the relationship between the predictive likelihoods p_{SLB} and p_{SLI}, it follows that p_{SLB} is also parameter invariant.

Attention will now be directed at finding a factorisation, as specified by (2.21), following the convention of Hinkley (1979). Let $f(x\,|\,\mathfrak{x})$ denote the posterior predictive density determined using the prior probability density function $f(\theta)$ of the parameter vector θ. Then, assuming $\mathcal{X}$ and X are dependent, it follows that

$$\begin{aligned} f(x\,|\,\mathfrak{x}) &= \frac{\int f(x,\mathfrak{x}\,|\,\theta)f(\theta)\,d\theta}{\int f(\mathfrak{x}\,|\,\theta)f(\theta)\,d\theta} \\ &= \frac{f(x,\mathfrak{x}\,|\,\theta)}{f(r\,|\,\theta)}\frac{\int f(r\,|\,\theta)f(\theta)\,d\theta}{\int f(\mathfrak{x}\,|\,\theta)f(\theta)\,d\theta} \\ &= p_{\mathrm{SLI}}(x\,;\mathfrak{x})\frac{f(r)}{f(\mathfrak{x})}. \end{aligned} \tag{2.29}$$

Hence a factorisation for the predictive likelihood p_{SLI} is specified.

Assume the sample matrix $\mathcal{X}$ and the predictand X are independent, with R, S, and T defined as at (2.22), where R has the same dimension as T. Let $f_{(S,R)}(s,r\,;\theta)$ denote the joint probability density function of S and R, and let $f_{(S,T)}(s,t\,;\theta)$ denote the joint probability density function of S and T. Let $\partial r/\partial t$ denote the Jacobian matrix of r with respect to t. Then it follows that

$$p_{\mathrm{SLH}}(x\,;\mathfrak{x}) = \frac{f_{\mathcal{X}}(\mathfrak{x}\,;\theta)f_{X}(x\,;\theta)}{f_{R}(r\,;\theta)\det(\partial r/\partial t)}.$$

Therefore, in a similar way to (2.29), it follows that

$$f(x\,|\,\mathfrak{x}) = p_{\mathrm{SLH}}(x\,;\mathfrak{x})\frac{\det(\partial r/\partial t)f(r)}{f(\mathfrak{x})}.$$

Hence a factorisation for the predictive likelihood p_{SLH} is specified.

As stated by Barndorff-Nielsen (1980), while using sufficiency for predictive likelihood is "expedient in eliminating the parameter the method seems lacking in primitive motivation." The contrived nature of sufficiency based predictive likelihood is not its only impediment. Notice that both definitions of predictive likelihood given by (2.24) and (2.28) rely on the existence of a minimal sufficient statistic. Pitcher (1957) and Landers and Rogge (1972) consider instances where no minimal sufficient statistic exits; therefore, the restrictive nature of any version of predictive likelihood based on sufficiency is exposed.

2.2.2 Approximate Predictive Likelihood

This subsection proposes versions of predictive likelihoods which do not depend on the notion of sufficiency; hence they will have wider applicability. Davison (1986, 1990) derives a predictive likelihood by using asymptotics to obviate the appearance of the unknown parameter vector θ as follows.

Recall that in a Bayesian framework, as previously discussed at the beginning of this chapter, predictive inference for the predictand X would be based on the posterior predictive density function $f(x \mid \mathfrak{x})$, given by (2.4), should a prior probability density $f(\theta)$ for the k-dimensional parameter vector θ be available. Assuming the sample $\mathcal{X}$ and the predictand X are dependent, the posterior predictive density function $f(x \mid \mathfrak{x})$ can be expressed in the form

$$f(x \mid \mathfrak{x}) = \frac{f(\mathfrak{x}, x)}{f(\mathfrak{x})}, \tag{2.30}$$

where

$$f(\mathfrak{x}, x) = \int f(\mathfrak{x}, x \mid \theta) f(\theta)\, d\theta, \qquad f(\mathfrak{x}) = \int f(\mathfrak{x} \mid \theta) f(\theta)\, d\theta, \tag{2.31}$$

$f(\mathfrak{x}, x \mid \theta) = f_{(\mathcal{X},X)}(\mathfrak{x}, x; \theta)$, and $f(\mathfrak{x} \mid \theta) = f_{\mathcal{X}}(\mathfrak{x}; \theta)$.

Suppose that the integrand for $f(\mathfrak{x})$, indexed by (2.31), has a well-defined mode as a function of θ and that $\log f(\mathfrak{x} \mid \theta)$ and $\log f(\theta)$ are twice continuously differentiable functions of θ. Then, using Laplace's method for integrals (de Bruijn, 1981; Bleistein and Handelsman, 1986; Barndorff-Nielsen and Cox, 1989) the integral denoted by $f(\mathfrak{x})$ can be expanded as an asymptotic series. Proceeding with this technique it may be shown that

$$f(\mathfrak{x}) = (2\pi)^{p/2} \det\{\hat{I}(\hat{\theta})\}^{-1/2} f_{\mathcal{X}}(\mathfrak{x}; \hat{\theta}) f(\hat{\theta}) \{1 + O_p(n^{-1})\}, \tag{2.32}$$

where $\hat{\theta}$ denotes the mode of the integrand for $f(\mathfrak{x})$ obtained as the solution of the set of equations $\partial \log f_{\mathcal{X}}(\mathfrak{x};\theta)/\partial\theta + \partial \log f(\theta)/\partial\theta = 0$, and $\hat{I}(\theta)$ denotes the $k \times k$ Hessian matrix of $-\log\{f_{\mathcal{X}}(\mathfrak{x};\theta)\} - \log\{f(\theta)\}$ (Magnus and Neudecker, 1999).

Additionally, suppose that the integrand for $f(x,\mathfrak{x})$, indexed by (2.31), has a well-defined mode as a function of θ and that $\log f(\mathfrak{x},x \mid \theta)$ is a twice continuously differentiable function of θ. Then, an asymptotic series can be determined for the integral $f(\mathfrak{x},x)$, again using Laplace's method for integrals, which takes the form

$$f(\mathfrak{x},x) = (2\pi)^{p/2} \det\{\hat{I}_X(\hat{\theta}_x)\}^{-1/2} f_{(\mathcal{X},X)}(\mathfrak{x},x;\hat{\theta}_x) f(\hat{\theta}_x)\{1 + O_p(n^{-1})\}, \tag{2.33}$$

where $\hat{\theta}_x$ is the mode of the integrand for $f(\mathfrak{x},x)$ obtained as the solution of the set of equations $\partial \log f_{(\mathcal{X},X)}(\mathfrak{x},x;\theta)/\partial\theta + \partial \log f(\theta)/\partial\theta = 0$, and $\hat{I}_X(\theta)$ denotes the $k \times k$ Hessian matrix of $-\log\{f_{(\mathcal{X},X)}(\mathfrak{x},x;\theta)\} - \log\{f(\theta)\}$. An asymptotic expansion of the posterior predictive density function $f(x \mid \mathfrak{x})$ is obtained by substituting (2.32) and (2.33) into (2.30), and is given by

$$f(x \mid \mathfrak{x}) = \frac{\det\{\hat{I}(\hat{\theta})\}^{1/2} f_{(\mathcal{X},X)}(\mathfrak{x},x;\hat{\theta}_x) f(\hat{\theta}_x)}{\det\{\hat{I}_X(\hat{\theta}_x)\}^{1/2} f_{\mathcal{X}}(\mathfrak{x};\hat{\theta}) f(\hat{\theta})}\{1 + O_p(n^{-1})\}. \tag{2.34}$$

Davison (1986) defines the approximate predictive likelihood for the predictand X as

$$p_{\mathrm{AP}}(x;\mathfrak{x}) = \frac{\det\{\hat{J}(\hat{\theta})\}^{1/2} f_{(\mathcal{X},X)}(\mathfrak{x},x;\hat{\theta}_x)}{\det\{\hat{J}_X(\hat{\theta}_x)\}^{1/2} f_{\mathcal{X}}(\mathfrak{x};\hat{\theta})}, \tag{2.35}$$

where $\hat{J}(\theta)$ and $\hat{J}_X(\theta)$ denote the observed information matrices (Barndorff-Nielsen and Cox, 1994, p. 25) which are defined as the Hessian matrices, with respect to the unknown parameter vector θ, of $-\log\{f_{\mathcal{X}}(\mathfrak{x};\theta)\}$ and $-\log\{f_{(\mathcal{X},X)}(\mathfrak{x},x;\theta)\}$, respectively. Note that (2.35) represents the leading term of the asymptotic series at (2.34) when the prior probability density of the parameter vector θ is constant for all $\theta \in \Theta$; hence the approximate predictive likelihood $p_{\mathrm{AP}}(x;\mathfrak{x})$ is obtained from the leading term of the asymptotic series determined by the expansion of

$$\frac{\int f_{(\mathcal{X},X)}(\mathfrak{x},x;\theta)\,d\theta}{\int f_{\mathcal{X}}(\mathfrak{x};\theta)\,d\theta}.$$

In a similar spirit to Davison (1986), two predictive functions are introduced by Butler (1989). The first is referred to as the approximate conditional predictive likelihood for the predictand X and is defined by

$$p_{\mathrm{APB}}(x;\mathfrak{x}) = \frac{f_{(\mathcal{X},X)}(\mathfrak{x},x;\hat{\theta}_x) \det\{\hat{J}_X(\hat{\theta}_x)\}^{1/2}}{\det\{J(\hat{\theta}_x)J(\hat{\theta})^{\top}\}^{1/2}}, \tag{2.36}$$

where $\hat{\theta}_x = \arg\max_\theta f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)$ and $J(\theta) = \partial^2 f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta)/\partial\theta\partial(\mathfrak{x}, x)^\top$. The second is referred to as the modified profile predictive likelihood for the predictand X, and, assuming the transformation taking $(X, \hat{\theta})$ to $(X, \hat{\theta}_X)$ is one-to-one, is defined by

$$p_{\mathrm{MP}}(x\,;\mathfrak{x}) = f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\hat{\theta}_x)\det\left\{\hat{J}_X(\hat{\theta}_x)\right\}^{-1/2}\det(K), \tag{2.37}$$

where $\hat{\theta} = \arg\max_\theta f_{\mathcal{X}}(\mathfrak{x}\,;\theta)$ and $K = \partial\hat{\theta}/\partial\hat{\theta}_x^\top$. The modified profile predictive likelihood is the predictive analogue of the so called modified profile likelihood of Barndorff-Nielsen (1983).

When the sample $\mathcal{X}$ and the predictand X have a joint probability density function

$$f_{(\mathcal{X},X)}(\mathfrak{x}, x\,;\theta) = \exp\left\{\theta^\top t(\mathfrak{x}, x) - c(\theta) - d(\mathfrak{x}, x)\right\}$$

belonging to a regular exponential family, for functions t, c and d, it can be shown (Butler, 1989, Lemma 1) that p_{APB} is a saddlepoint approximation of p_{SLB} (see (2.28)) since $p_{\mathrm{APB}}(x\,;\mathfrak{x}) \propto p_{\mathrm{SLB}}(x\,;\mathfrak{x})\{1 + O(n^{-1})\}$, where the $O(n^{-1})$ term depends on x. Additionally, suppose the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denotes n independent and identically distributed random vectors from a population with probability density function $f_X(x\,;\theta) = \exp\left\{\theta^\top q(x) - c(\theta) - d(x)\right\}$ belonging to a regular exponential family, for functions q, c and d. Then, if the predictand X is independent of the sample $\mathcal{X}$ but drawn from the same population, p_{MP} (see (2.37)) is the leading term from the asymptotic expansion, derived using Laplace's method of integrals, of the posterior predictive density function $f(x \mid \mathfrak{x})$ when the prior probability density function of the parameter vector θ is given by $f(\theta) \propto \det\left\{\hat{J}_X(\hat{\theta}_x)\right\}$ or is constant on $E(X) = \partial c/\partial\theta$.

Both predictive likelihoods p_{AP} and p_{MP} are invariant under scale changes of the predictand X, while p_{APB} and p_{MP} are parameter invariant. It should also be stressed that when $\hat{\theta}$ is not a function of $(x, \hat{\theta}_x)$ (Bjørnstad, 1990, Example 4) the predictive likelihood p_{MP} cannot be constructed for predictive inference. Of the predictive likelihoods considered in this section, p_{AP} and p_{APB} are of wide applicability even though they have complimentary invariance properties.

2.3 Predictive Functions

This section proposes predictive functions for the predictand X which may be used for predictive inference. Whereas the construction of predictive likelihoods in the previous section was somewhat contrived, this section proposes predictive functions using an axiomatic procedure.

Let Y denote a random vector with probability density function given by $f_Y(y;\theta)$ where the unknown parameter vector $\theta \in \Theta$ is a member of the parameter space Θ with y denoting a realisation of Y. Barnard (1949) assesses the verisimilitude of the pair (y,θ) via an absolute odds-function $\Psi(y,\theta)$. Three absolute odds-functions are given by

$$\Psi_1(y,\theta) = f_Y(y;\theta), \qquad \Psi_2(y,\theta) = \frac{f_Y(y;\theta)}{\sup_y f_Y(y;\theta)}, \qquad \text{and} \qquad \Psi_3(y,\theta) = \frac{f_Y(y;\theta)}{\sup_\theta f_Y(y;\theta)}.$$

Assuming the sample $\mathcal{X}$ and the predictand X are independent, Mathiasen (1979) assesses the verisimilitude of the triple $(\mathfrak{x},x,\theta)$ using the function $\Upsilon(\mathfrak{x},x,\theta) = \Psi(\mathfrak{x},\theta)\Psi(x,\theta)$. In general, a generic prediction function, $\check{p}(x;\mathfrak{x})$, is then defined by setting

$$\check{p}(x;\mathfrak{x}) = \sup_\theta \Upsilon(\mathfrak{x},x,\theta).$$

When the absolute odds-function Ψ is given by Ψ_1, Ψ_2, and Ψ_3, the derived predictive functions are

$$p_{\mathrm{L}}(x;\mathfrak{x}) = \sup_\theta f_X(x;\theta) f_{\mathcal{X}}(\mathfrak{x};\theta),$$
$$p_{\mathrm{P}}(x;\mathfrak{x}) = \sup_\theta \left\{ \frac{f_X(x;\theta)}{\sup_x f_X(x;\theta)} \frac{f_{\mathcal{X}}(\mathfrak{x};\theta)}{\sup_{\mathfrak{x}} f_{\mathcal{X}}(\mathfrak{x};\theta)} \right\},$$

and

$$p_{\mathrm{F}}(x;\mathfrak{x}) = \sup_\theta \left\{ \frac{f_X(x;\theta)}{\sup_\theta f_X(x;\theta)} \frac{f_{\mathcal{X}}(\mathfrak{x};\theta)}{\sup_\theta f_{\mathcal{X}}(\mathfrak{x};\theta)} \right\}, \tag{2.38}$$

respectively.

Using the axiomatic framework of Mathiasen (1979), along with a slightly different approach for predictive function construction, Barndorff-Nielsen (1978) considers a whole class of predictive functions of which the likelihood predictive function for the predictand X, defined by

$$p_{\mathrm{LP}}(x;\mathfrak{x}) = \sup_\theta \left\{ f_{\mathcal{X}}(\mathfrak{x};\theta) \frac{f_X(x;\theta)}{\sup_x f_X(x;\theta)} \right\},$$

is a member. The distributional form of p_{LP} was shown by Barndorff-Nielsen (1980) to be concordant with classical solutions, for example, those obtained from pivotal transformation considerations.

When the sample $\mathcal{X}$ and the predictand X are dependent, the four predictive functions p_{L}, p_{P},

p_{F}, and p_{LP} generalise to

$$p_{\mathrm{L}}(x\,;\mathfrak{x}) = \sup_{\theta} f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta),$$

$$p_{\mathrm{P}}(x\,;\mathfrak{x}) = \sup_{\theta}\left\{\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{\sup_{\mathfrak{x}} f_{\mathcal{X}}(\mathfrak{x}\,;\theta)\sup_{x|\mathfrak{x}} f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x})}\right\},$$

$$p_{\mathrm{F}}(x\,;\mathfrak{x}) = \sup_{\theta}\left\{\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{\sup_{\theta} f_{\mathcal{X}}(\mathfrak{x}\,;\theta)\sup_{\theta} f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x})}\right\},$$

and

$$p_{\mathrm{LF}}(x\,;\mathfrak{x}) = \sup_{\theta}\left\{\frac{f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)}{\sup_{x|\mathfrak{x}} f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x})}\right\}.$$

Set $\hat{\theta}_x = \arg\max_{\theta} f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)$; then the predictive function p_{L} may be represented as $p_{\mathrm{L}}(x\,;\mathfrak{x}) = f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\hat{\theta}_x)$. From this observation, the intrinsic involvement of the predictive function p_{L} in the predictive likelihoods given by p_{AP}, p_{APB} and p_{MP}, see (2.35), (2.36), and (2.37), respectively, is evident. It should be noted that p_{L} is scale and parameter invariant and so is desirable as a predictive function per se.

Not all the former predictive functions can be used for predictive inference for all parametric populations. For example, let the sample $\mathcal{X}$ denote n independent and identically distributed random variables from a normal population with unknown mean μ and variance σ^2, and suppose the predictand X is a random variable which is independent of $\mathcal{X}$ but drawn from the same population. Then the predictive function p_{F} given by (2.38) is not defined because $\sup_{\theta} f_X(x\,;\theta) = \infty$. Note also that in the case when p_{F} is well defined, p_{L} is also well defined, and thus acts as a competitor.

2.4 Predictive Densities

This section proposes the use of two types of predictive density for predictive inference. The first type is derived by directly constructing an estimator of $f_{X|\mathcal{X}}(x\,;\theta_0\,|\,\mathfrak{x})$, the true conditional probability density function of the predictand X given the sample $\mathcal{X} = \mathfrak{x}$, using the sample only. The second type is derived from retrospection; asymptotic considerations of the upper end-point of a nominal one-sided α-level prediction interval imply an appropriate generating density which is based on the sample only. These approaches are delineated as follows.

Let $f_{(\mathcal{X},X)}(\mathfrak{x},x\,;\theta)$ denote the joint probability density function of the sample $\mathcal{X}$ and the predictand X, where θ is the unknown parameter vector with true value θ_0. Additionally, denote the

maximum likelihood estimate of θ, based on the sample $\mathcal{X}$, by $\hat{\theta} = \arg\max_{\theta} f_{\mathcal{X}}(\mathfrak{x}\,;\theta)$. An estimator of $f_{X|\mathcal{X}}(x\,;\theta_0\,|\,\mathfrak{x})$ (Rao, 1975), referred to as the estimative conditional predictive density for the predictand X and defined by

$$p_{\mathrm{E}}(x\,;\mathfrak{x}) = f_{X|\mathcal{X}}(x\,;\hat{\theta}\,|\,\mathfrak{x}), \tag{2.39}$$

is obtained by replacing the unknown parameter vector θ by $\hat{\theta}$.

If $\mathcal{X}$ and X are independent, (2.39) is referred to as the estimative predictive density for the predictand X, specialises to

$$p_{\mathrm{E}}(x\,;\mathfrak{x}) = f_X(x\,;\hat{\theta}), \tag{2.40}$$

and represents an estimator of $f_X(x\,;\theta_0)$, the probability density function of the predictand X, where the unknown parameter vector θ is replaced by $\hat{\theta}$.

Kalbfleisch and Sprott (1970, 1972), Aitchison and Dunsmore (1975), Butler (1986), and Bjørnstad (1990) contend that when the sample size n is small or the relative dimension of θ is large, p_{E} will be a poor choice for predictive inference. An approach which acknowledges the random characteristic of $\hat{\theta}$, the plug-in estimate employed for p_{E}, is as follows.

Denote by $f_{\hat{\theta}}(\cdot\,;\theta)$ the probability density function of the maximum likelihood estimate $\hat{\theta}$ based on the sample $\mathcal{X}$, which is assumed to be independent of the predictand X. Then, as an estimator for $f_X(x\,;\theta_0)$, Harris (1989) proposes the parametric bootstrap predictive density for the predictand X, which is given by

$$p_{\mathrm{PB}}(x\,;\mathfrak{x}) = \int f_X(x\,;\vartheta) f_{\hat{\theta}}(\vartheta\,;\hat{\theta})\, d\vartheta, \tag{2.41}$$

or equivalently,

$$p_{\mathrm{PB}}(x\,;\mathfrak{x}) = \left[\int f_X\{x\,;\hat{\theta}(\mathfrak{x})\} f_{\mathcal{X}}(\mathfrak{x}\,;\theta)\, d\mathfrak{x} \right]_{\theta=\hat{\theta}},$$

and can be verbalised as follows: the parametric bootstrap predictive density for the predictand X is expressed as the expected value of p_{E} at (2.40) in which the unknown parameter vector θ is replaced by the maximum likelihood estimate $\hat{\theta}$.

While p_{PB} is parameter and scale invariant it should be noted that (2.41) is rarely obtained in closed form; recourse can be had to numerical computation. As a computational expedient, Vidoni

(1995) considers an approximation to (2.41) that is obtained through asymptotic arguments and delineated as follows.

Suppose the maximum likelihood estimate $\hat{\theta} = \arg\max_{\theta} f_{\mathcal{X}}(\mathfrak{x};\theta)$ conjoined with an ancillary statistic A, constitutes a sufficient statistic for the sample $\mathcal{X}$; that is, A has a distribution which does not depend on the unknown parameter vector θ (Barndorff-Nielsen and Cox, 1994, Section 2.5). Therefore, without loss of generality, the sample $\mathcal{X}$ is represented by $(\hat{\theta}, A)$ and the log-likelihood $l(\theta;\mathfrak{x}) = \log f_{\mathcal{X}}(\mathfrak{x};\theta)$ may be written as $l(\theta;\hat{\theta},a)$, where a denotes a realisation of the ancillary statistic A. This follows from the factorisation theorem (Barndorff-Nielsen and Cox, 1994, Section 2.3): essentially, a necessary and sufficient condition for $(\hat{\theta}, A)$ to be sufficient for θ is that for all $\mathfrak{x}$ and $\theta \in \Theta$,

$$f_{\mathcal{X}}(\mathfrak{x};\theta) = g(\hat{\theta},a;\theta)h(\mathfrak{x})$$

for some functions $g(\hat{\theta},a;\theta)$ and $h(\mathfrak{x})$. Without loss of generality, $g(\hat{\theta},a;\theta) = f_{(\hat{\theta},A)}(\hat{\theta},a;\theta)$ the joint probability density function of $\hat{\theta}$ and A. To advance calculations, it is advisable, (Barndorff-Nielsen and Cox, 1994) to replace $f_{\hat{\theta}}(\hat{\theta};\theta)$, the probability density function of the maximum likelihood estimator $\hat{\theta}$ employed in (2.41), by $f_{\hat{\theta}}(\hat{\theta};\theta \mid a)$, the conditional probability density function of $\hat{\theta}$ given the ancillary statistic $A = a$, to obtain

$$p_{\mathrm{PB}}(x;\mathfrak{x}) = \int f_X(x;\vartheta) f_{\hat{\theta}}(\vartheta;\hat{\theta} \mid a)\, d\vartheta, \tag{2.42}$$

the conditional parametric bootstrap predictive density for the predictand X. Properties of (2.42) which hold conditionally on A will also hold unconditionally via the tower property of conditional expectation (see (2.12)).

Assume that the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^{\top}$ denotes n independent and identically distributed random variables drawn from a population with probability density function $f_X(x;\theta)$, where θ is the unknown parameter vector, and the predictand X is independent of $\mathcal{X}$ but drawn from the same population. Barndorff-Nielsen (1983) (Barndorff-Nielsen and Cox, 1994, Chapter 6) considers as an approximation to $f_{\hat{\theta}}(\hat{\theta};\theta \mid a)$ the probability density function $f^{*}_{\hat{\theta}}(\hat{\theta};\theta \mid a)$ defined by

$$f^{*}_{\hat{\theta}}(\hat{\theta};\theta \mid a) = c(\theta,a) \det(\hat{j})^{1/2} \exp\left\{l(\theta;\hat{\theta},a) - l(\hat{\theta};\hat{\theta},a)\right\}, \tag{2.43}$$

where $c(\theta,a)$ is a normalising constant selected such that

$$\int f^{*}_{\hat{\theta}}(\hat{\theta};\theta \mid a)\, d\hat{\theta} = 1$$

and $\hat{j} = -\nabla_\theta \nabla_\theta^\top l(\theta\,;\hat{\theta},a)|_{\theta=\hat{\theta}}$ denotes the observed information matrix evaluated at $\theta = \hat{\theta}$ and considered as a function of $\hat{\theta}$.

While $f_{\hat{\theta}}^*$ is sample and parameter invariant, it is also identical to $f_{\hat{\theta}}$ for a range of models (Barndorff-Nielsen and Cox, 1994, p. 175). Under appropriate regularity conditions, it can be shown that

$$f_{\hat{\theta}}(\hat{\theta}\,;\theta \,|\, a) = f_{\hat{\theta}}^*(\hat{\theta}\,;\theta \,|\, a)\{1 + O(n^{-3/2})\}$$

whenever $|\hat{\theta} - \theta| \leq cn^{-1/2}$, for some nonnegative constant c.

By replacing $f_{\hat{\theta}}$ by $f_{\hat{\theta}}^*$ in (2.42), Vidoni (1995) proposed the $f_{\hat{\theta}}^*$-conditional parametric bootstrap predictive density, defined as

$$p_{\mathrm{PB}}^*(x\,;\mathfrak{x}) = \int f_X(x\,;\vartheta) f_{\hat{\theta}}^*(\vartheta\,;\hat{\theta} \,|\, a)\, d\vartheta, \tag{2.44}$$

which, under appropriate regularity conditions, satisfies

$$p_{\mathrm{PB}}(x\,;\mathfrak{x}) = p_{\mathrm{PB}}^*(x\,;\mathfrak{x})\{1 + O(n^{-3/2})\}.$$

Under the regularity conditions given in Barndorff-Nielsen and Cox (1989, Chapter 6), (2.44) can by approximated using Laplace's method of integrals (de Bruijn, 1981; Bleistein and Handelsman, 1986; Barndorff-Nielsen and Cox, 1989; Tierney, Kass, and Kadane, 1989) to yield the approximate $f_{\hat{\theta}}^*$-conditional parametric bootstrap predictive density for the predictand X defined as

$$p_{\mathrm{APB}}^*(x\,;\mathfrak{x}) = p_{\mathrm{E}}(x\,;\mathfrak{x})\{1 + \tfrac{1}{2}G(x\,;\hat{\theta},a)\},$$

where

$$\begin{aligned} G(x\,;\theta,a) = \{&l_X(\theta\,;x)_{ij;}\, + l_X(\theta\,;x)_{i;}\, l_X(\theta\,;x)_{j;}\, + 2l_X(\theta\,;x)_{i;}\, \check{s}_j(\theta\,;a)\} r^{ij}(\theta\,;\theta,a) \\ &- l_X(\theta\,;x)_{i;}\, r_{jkl}(\theta\,;\theta,a) r^{jk}(\theta\,;\theta,a) r^{li}(\theta\,;\theta,a). \end{aligned}$$

Here $l_X(\theta\,;x) = \log f_X(x\,;\theta)$, $r(\theta\,;\hat{\theta},a) = l(\theta\,;\theta,a) - l(\hat{\theta}\,;\theta,a)$, $s(\theta\,;a) = \det(\hat{j})^{1/2}|_{\hat{\theta}=\theta}$, $\check{s}_j(\theta\,;a) = s_{/j}(\theta\,;a)/s(\theta\,;a)$, and $r^{ij}(\theta\,;\theta,a)$ denotes the (i,j)-th element from the inverse of the matrix $[r_{ij}(\theta\,;\theta,a)]$ with $l_X(\theta\,;x)_{R;} = l_X(\theta\,;x)_{/R;}$, $s_{/R}(\theta\,;a)$, and $r_R(\theta\,;\hat{\theta},a) = r_{/R}(\theta\,;\hat{\theta},a)$ corresponding to the partial derivatives with respect to the components of θ with indices in the index set R. Under appropriate regularity conditions, it can be shown that

$$p_{\mathrm{PB}}(x\,;\mathfrak{x}) = p_{\mathrm{APB}}^*(x\,;\mathfrak{x})\{1 + O(n^{-3/2})\}.$$

Therefore, while requiring a higher level of technical sophistication, p^*_{APB} provides a credible closed form alternative to p_{PB}.

The remainder of this section will delineate the second type of predictive density in which it will be assumed that the predictand X is a random variable whose dependency on the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ is represented by $f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x})$, the conditional probability density function of X given $\mathcal{X} = \mathfrak{x}$. The rationale behind the derivation of this predictive density progresses by first considering an exact one-sided α-level prediction interval $\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{q}_\alpha]$, where $\hat{q}_\alpha = \hat{q}_\alpha(\mathcal{X})$ depends on the sample $\mathcal{X}$ and $\alpha \in (0,1)$ only, and satisfies

$$\begin{aligned}\alpha &= P(X \in \widehat{\mathcal{I}}_\alpha) \\ &= E\left\{ \int_{-\infty}^{\hat{q}_\alpha(\mathcal{X})} f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathcal{X})\,dx \right\} \\ &= \int F_{X|\mathcal{X}}(\hat{q}_\alpha(\mathfrak{x})\,;\theta\,|\,\mathfrak{x}) f_{\mathcal{X}}(\mathfrak{x}\,;\theta)\,d\mathfrak{x}, \end{aligned} \tag{2.45}$$

where $F_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x}) = \int_{-\infty}^{x} f_{X|\mathcal{X}}(y\,;\theta\,|\,\mathfrak{x})\,dy$ denotes the conditional probability distribution function of the predictand X given the sample $\mathcal{X} = \mathfrak{x}$.

For any n, an explicit expression for $\hat{q}_\alpha$, which depends only on the sample $\mathcal{X}$ and α, can be constructed whenever a pivotal transformation (see Section 2.1) for the predictand X and the sample $\mathcal{X}$ exists; in general, an asymptotic expansion for $\hat{q}_\alpha$ can be derived.

Let $(\hat{\theta}, A)$ denote a function of the sample $\mathcal{X}$ only. Then, essentially, $(\hat{\theta}, A)$ is prediction-sufficient for the predictand X provided it is transitive, i.e.

$$f_{X|\mathcal{X}}(x\,;\theta\,|\,\mathfrak{x}) = f_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta}, a),$$

and sufficient for $\mathcal{X}$ with respect to parametric inference on θ (see Barndorff-Nielsen and Cox, 1996, Section 3, for references), where $f_{X|(\hat{\theta},A)}$ denotes the probability density function of X given $(\hat{\theta}, A)$ and a denotes a realisation of A. Additionally, under appropriate regularity conditions, $(\hat{\theta}, A)$ is prediction-sufficient for the predictand X if and only if $(\hat{\theta}, A)$ is sufficient for the class of all conditional distributions of $\mathcal{X}$ given X considered as a parametric family with parameters θ and x.

Assuming $(\hat{\theta}, A)$ is a prediction-sufficient reduction of $\mathcal{X}$ with ancillary statistic A select $\hat{q}_\alpha = \hat{q}_\alpha(\hat{\theta}, A)$, which depends on the sample through the maximum likelihood estimator $\hat{\theta}$ and the

ancillary statistic A, to satisfy

$$\begin{aligned}\alpha &= P(X \in \widehat{\mathcal{I}}_\alpha \,|A) \\ &= \int F_{X|(\hat\theta,A)}\{\hat q_\alpha(\vartheta,a)\,;\theta\,|\,\vartheta,a\} f_{\hat\theta|A}(\vartheta\,;\theta\,|a)\,d\vartheta,\end{aligned} \tag{2.46}$$

where $F_{X|(\hat\theta,A)}(x\,;\theta\,|\,\vartheta,a)$ denotes the conditional probability distribution function of the predictand X given $(\hat\theta, A) = (\vartheta, a)$. Using the tower property of conditional expectation, $\hat q_\alpha(\hat\theta, A)$ satisfies the unconditional version of (2.46),

$$\alpha = P(X \in \widehat{\mathcal{I}}_\alpha) = P\{X \le \hat q_\alpha(\hat\theta, A)\},$$

analogous to (2.45).

According to Barndorff-Nielsen and Cox (1996), and under appropriate regularity conditions,

$$\hat q_\alpha(\theta, a) = \hat q_{\alpha,2}(\theta, a) + O(n^{-3/2}), \tag{2.47}$$

in which $\hat q_{\alpha,2}(\theta, a)$ satisfies

$$F_{X|(\hat\theta,A)}\{\hat q_{\alpha,2}(\theta,a)\,;\theta\,|\,\theta,a\} = \alpha - R(\theta,a),$$

where

$$R(\theta,a) = Q(\hat q_{\alpha,1}(\theta,a)\,;\theta\,|\,\theta,a)$$

and $\hat q_{\alpha,1}(\theta, a)$ satisfies

$$F_{X|(\hat\theta,A)}\{\hat q_{\alpha,1}(\theta,a)\,;\theta\,|\,\theta,a\} = \alpha.$$

Since $(\hat\theta, A)$ is prediction-sufficient it is also sufficient, and, via the factorisation theorem (Barndorff-Nielsen and Cox, 1994, Section 2.3), it can be assumed, without loss of generality, that the log-likelihood function for θ based on the sample $\mathcal{X}$ depends on the sample $\mathcal{X}$ through $(\hat\theta, A)$ only. To express this conclusion, write $l(\theta\,;\hat\theta, a)$ for the log-likelihood function for θ based on the sample, and define

$$\check l_{R_1;\,R_2} = l_{/R_1;\,R_2}(\theta\,;\hat\theta,a)|_{\hat\theta=\theta}, \tag{2.48}$$

the partial derivative of $l(\theta\,;\hat\theta, a)$ with respect to the components of θ and $\hat\theta$ given by the index sets R_1 and R_2 respectively, evaluated at $\hat\theta = \theta$. (For precedents for such notation see Barndorff-Nielsen

and Cox, 1989, Chapter 5, and Barndorff-Nielsen and Cox, 1994, Chapter 5.) Let $\check{j}^{ij}$ denote the (i,j)-th element from the inverse of the matrix $[-\check{l}_{ij;}]$, then,

$$Q(x;\theta,a) = \tfrac{1}{2}(\check{h}_{rs} - \check{h}_t \check{l}_{u;\,rs}\check{j}^{tu})\check{j}^{rs}, \tag{2.49}$$

where $\check{h}_r = -F_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,r;}$,

$$\begin{aligned}\check{h}_{rs} = f_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)^{-1}\{&F_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,r;}\\ &+F_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,;\,r}\}f_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,s;}\,[2]\\ &-F_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,rs;} - F_{X|(\hat{\theta},A)}(x;\theta\,|\,\theta,a)_{/;\,r;\,s}[2],\end{aligned}$$

and, $f_{X|(\hat{\theta},A)}(x;\theta\,|\,\hat{\theta},a)_{/R_1;\,R_2;\,R_3}$ and $F_{X|(\hat{\theta},A)}(x;\theta\,|\,\hat{\theta},a)_{/R_1;\,R_2;\,R_3}$ represent the partial derivative of $f_{X|(\hat{\theta},A)}(x;\theta\,|\,\hat{\theta},a)$ and $F_{X|(\hat{\theta},A)}(x;\theta\,|\,\hat{\theta},a)$, respectively, with respect to the components of x, θ and $\hat{\theta}$ given by the index sets R_1, R_2, and R_3, respectively.

For any index set R (Barndorff-Nielsen and Cox, 1994, p. 148) it can be shown that

$$\check{l}_{R;} + \sum_{R/2} \check{l}_{R_1;\,R_2} = 0. \tag{2.50}$$

Therefore, when $|R| = 3$, (2.50) specialises to give

$$\check{l}_{rsu;} + \check{l}_{u;\,rs} + \check{l}_{ru;\,s} + \check{l}_{su;\,r} = 0. \tag{2.51}$$

Rearranging $\check{l}_{u;\,rs}$ from (2.51) and substituting into (2.49), it follows that

$$Q(x;\theta,a) = \tfrac{1}{2}\{\check{h}_{rs} + \check{h}_t(\check{l}_{rsu;} + \check{l}_{ru;\,s} + \check{l}_{su;\,r})\check{j}^{tu}\}\check{j}^{rs}. \tag{2.52}$$

Alternatively, without adjusting the order at (2.47) it is possible to replace the quantities $\check{j}^{tu}$ and $\check{l}_{u;\,rs}$ of (2.49) by i^{tu} and $i_{u,rs} + i_{r,s,u}$ respectively (see Barndorff-Nielsen and Cox, 1994, Section 5.5), obtaining,

$$\tilde{Q}(x;\theta,a) = \tfrac{1}{2}\{\check{h}_{rs} + \check{h}_t(i_{u,rs} + i_{u,r,s})i^{tu}\}i^{rs}, \tag{2.53}$$

where $i_{u,rs} = E\{l_{/u;}\,(\theta;\hat{\theta},a)l_{/rs;}\,(\theta;\hat{\theta},a)\}$, $i_{r,s,u} = E\{l_{/r;}\,(\theta;\hat{\theta},a)l_{/s;}\,(\theta;\hat{\theta},a)l_{/u;}\,(\theta;\hat{\theta},a)\}$, and i^{tu} denotes the (t,u)-th element from the inverse of the matrix $[-E\{l_{/tu;}\,(\theta;\hat{\theta},a)\}]$.

When the sample $\mathcal{X}$ and the predictand X are independent the conditional distribution function $F_{X|(\hat{\theta},A)}(x;\theta\,|\,\hat{\theta},a) = F_X(x;\theta)$ does not depend on the sample through the maximum likelihood

estimator $\hat{\theta}$ or the ancillary statistic A, where F_X denotes the probability distribution function of the predictand X. Owing to this reduction, $\check{h}_r$ and $\check{h}_{rs}$ simplify to

$$\check{h}_r = -F_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)_{/;\,r;}$$

and

$$\check{h}_{rs} = f_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)^{-1}F_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)_{/;\,r;}\,f_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)_{/;\,s;}\,[2] \\ - F_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)_{/;\,rs;}\,,$$

with $\tilde{Q}$, given by (2.53), not depending on the ancillary statistic $A = a$.

From the asymptotic expansion for $\hat{q}_\alpha(\hat{\theta},A)$ given by (2.47), where θ is replaced by $\hat{\theta}$ and $Q(x\,;\theta,a)$ is defined by (2.49), (2.52) or (2.53), consider the retrospective construction of a predictive density $\check{p}(x\,;\mathfrak{x})$ such that

$$\int_{-\infty}^{\hat{q}_\alpha(\hat{\theta},a)} \check{p}(x\,;\mathfrak{x})\,dx = \alpha, \tag{2.54}$$

$\check{p}(x\,;\mathfrak{x}) \geq 0$, and $\int \check{p}(x\,;\mathfrak{x})\,dx = 1$, where $\hat{\theta} = \hat{\theta}(\mathfrak{x})$ and $a = a(\mathfrak{x})$ are explicitly dependent on the sample. To identify $\check{p}$, proceed by differentiating (2.54) with respect to α, to obtain

$$\check{p}\{\hat{q}_\alpha(\hat{\theta},a)\,;\hat{\theta},a\} = \{\hat{q}_\alpha(\hat{\theta},a)_{/\alpha}\}^{-1}.$$

Employing implicit differentiation and Taylor expansion in conjunction with the former expression, it was shown by Barndorff-Nielsen and Cox (1996, Section 6) that, under appropriate regularity conditions,

$$\check{p}(x\,;\mathfrak{x}) = p_{\mathrm{QI}}(x\,;\mathfrak{x}) + O(n^{-3/2}),$$

where $p_{\mathrm{QI}}(x\,;\mathfrak{x})$ is proposed by Barndorff-Nielsen and Cox (1996), referred to as the quantile inferred conditional predictive density for the predictand X, and is defined by

$$p_{\mathrm{QI}}(x\,;\mathfrak{x}) = \{1 + r_{/x}(x\,;\hat{\theta},a)\}f_{X|(\hat{\theta},A)}\{x + r(x\,;\hat{\theta},a)\,;\hat{\theta}\,|\,\hat{\theta},a\} \tag{2.55}$$

where

$$r(x\,;\theta,a) = \frac{Q(x\,;\theta,a)}{f_{X|(\hat{\theta},A)}(x\,;\theta\,|\,\hat{\theta},a)},$$

$\hat{\theta} = \arg\max_\theta f_{\mathcal{X}}(\mathfrak{x};\theta)$ denotes the maximum likelihood estimator of θ based on the sample $\mathcal{X}$, $r_{/x}(x;\hat{\theta},a)$ denotes the partial derivative of r with respect to x, and $Q(x;\theta)$ is defined by (2.49), (2.52) or (2.53).

Considerable technical expertise is generally required for evaluation of (2.55). However, p_{QI} is parameter, sample, and predictand invariant with $\int p_{\mathrm{QI}}(x;\mathfrak{x}) = 1$, and so augurs well for prediction interval construction.

2.5 Coverage Error and Calibration

As previously elaborated, an exact α-level prediction region can be constructed whenever a pivotal transformation for the sample and predictand exists. However, a nominal α-level prediction region can be constructed via a predictive likelihood, function or density (see Sections 2.2, 2.3 and 2.4, respectively). In the latter case it is natural for the relative performances of conformed prediction regions to be based on coverage error; a smaller absolute coverage error is favoured.

This section proceeds by elucidating the coverage error properties for, firstly, a nominal one-sided α-level prediction interval constructed via an estimative-type predictive density, and secondly, two nominal α-level prediction regions constructed via the estimative predictive density and the approximate predictive likelihood, respectively. Two types of analytic calibration — additive and level — are proposed for the former prediction interval, whereas parametric bootstrap level calibration is proposed for the latter prediction regions. In both instances, calibration is effective at reducing coverage error. (In contrast the conditional form of coverage error, conditional on the sample, usually cannot exceed $O_p(n^{-1/2})$, irrespective of calibration occurring (Beran, 1992).) A quantile-type predictive density is also retrospectively determined from the upper end-point of an additive-calibrated one-sided prediction interval constructed via the estimative predictive density. The section concludes with an evaluation of the computational burden for prediction region construction, and a review of an alternative form of empirical calibration.

Suppose the predictand X is a random variable which is independent of the sample $\mathcal{X}$, and let $F_X(x;\theta) = \int_{-\infty}^{x} f_X(y;\theta)\,dy$ denote the probability distribution function of X. Define the theoretical one-sided α-level prediction interval $\mathcal{I}_\alpha = (-\infty, q_\alpha(\theta)]$, where $q_\alpha(\theta)$ is the α-th quantile of $F_X(x;\theta)$. Then,

$$P(X \in \mathcal{I}_\alpha) = E[F_X\{q_\alpha(\theta);\theta\}] = \alpha.$$

Suppose a $\sqrt{n}$-consistent estimator $\tilde{\theta}$, a function of $\mathcal{X}$, of θ is available. For example, the maximum likelihood estimator $\hat{\theta} = \arg\max_{\theta} f_{\mathcal{X}}(\mathfrak{x};\theta)$ is a convenient choice for $\tilde{\theta}$ since it has well-known theoretical properties. Consider the nominal one-sided α-level prediction interval,

$$\widehat{\mathcal{I}}_{\alpha} = (-\infty, q_{\alpha}(\tilde{\theta})],$$

obtained by replacing θ by $\tilde{\theta}$ in $\mathcal{I}_{\alpha}$. Equivalently, $\widehat{\mathcal{I}}_{\alpha}$ can be constructed by calculating the α-th quantile of the probability distribution function $F_X(x;\tilde{\theta}) = \int_{-\infty}^{x} p_{\mathrm{ET}}(y;\mathfrak{x})\,dy$, where

$$p_{\mathrm{ET}}(x;\mathfrak{x}) = f_X(x;\tilde{\theta})$$

denotes an estimative-type predictive density for the predictand X. Note that $p_{\mathrm{ET}}(x;\mathfrak{x}) = p_{\mathrm{E}}(x;\mathfrak{x})$ (see (2.40)) whenever $\tilde{\theta}$ is identic to $\hat{\theta}$.

Under appropriate regularity conditions, by using the tower property (see (2.12)) with $\mathcal{G}_1 = \{\emptyset, \Omega\}$, $\mathcal{G}_2 = \sigma(\mathcal{X})$, and Taylor expansion, Cox (1973) and Barndorff-Nielsen and Cox (1994, Section 9.4) have shown that

$$P(X \in \widehat{\mathcal{I}}_{\alpha}) = E\{G(\tilde{\theta};\theta)\} = \alpha + d(\theta)/n + O(n^{-3/2}) \tag{2.56}$$

whenever $E(\tilde{\theta}) = \theta + b(\theta)/n + o(n^{-1})$ and $\mathrm{Var}(\tilde{\theta}) = c(\theta)/n + o(n^{-1})$, where

$$d(\theta) = \Big[b(\theta)^{\top}\nabla_{\tilde{\theta}} G(\tilde{\theta};\theta) + \tfrac{1}{2}\mathrm{tr}\{c(\theta) H_{\tilde{\theta}} G(\tilde{\theta};\theta)\}\Big]_{\tilde{\theta}=\theta},$$

$G(\tilde{\theta};\theta) = F_X\{q_{\alpha}(\tilde{\theta});\theta\}$, $\nabla_{\tilde{\theta}} G(\tilde{\theta};\theta)$ and $H_{\tilde{\theta}} G(\tilde{\theta};\theta) = \nabla_{\tilde{\theta}} \nabla_{\tilde{\theta}}^{\top} G(\tilde{\theta};\theta)$ denotes the gradient vector and Hessian matrix, respectively, of $G(\tilde{\theta};\theta)$ with respect to the vector $\tilde{\theta}$ (Magnus and Neudecker, 1999, pp. 87, 100), and, $b(\theta)$ and $c(\theta)$ denote a $1 \times k$ vector and a $k \times k$ matrix, respectively. When $\tilde{\theta}$ is identical to $\hat{\theta}$ it can be shown that $c(\theta) = -n^{-1} E\{H_{\theta} \log f_{\mathcal{X}}(\mathfrak{x};\theta)\}$, the expected information matrix per observation (Barndorff-Nielsen and Cox, 1994, Section 3).

Paraphrasing (2.56), the coverage error associated with the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\alpha}$ is $O(n^{-1})$. In an attempt to achieve further reduction in coverage error, an additive and level calibrated version of $\widehat{\mathcal{I}}_{\alpha}$, given by

$$\widehat{\mathcal{I}}_{\mathrm{AC},\alpha} = (-\infty, q_{\alpha}(\tilde{\theta}) + r(\alpha;\tilde{\theta})]$$

and

$$\widehat{\mathcal{I}}_{\mathrm{LC},\alpha} = (-\infty, q_{\beta(\alpha;\tilde{\theta})}(\tilde{\theta})],$$

are proposed by Barndorff-Nielsen and Cox (1994, Section 9.4) to extirpate the $O(n^{-1})$ term from the expansion at (2.56). Under appropriate regularity conditions and from direct analytic investigation, it was shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{AC},\alpha}) = \alpha + O(n^{-3/2}) \quad \text{and} \quad P(X \in \widehat{\mathcal{I}}_{\mathrm{LC},\alpha}) = \alpha + O(n^{-3/2}),$$

whenever

$$r(\alpha\,;\tilde{\theta}) = -n^{-1}d(\tilde{\theta})/f_X\{q_\alpha(\tilde{\theta})\,;\tilde{\theta}\} \quad \text{and} \quad \beta(\alpha\,;\tilde{\theta}) = \alpha - d(\tilde{\theta})/n.$$

Therefore $\widehat{\mathcal{I}}_{\mathrm{AC},\alpha}$ or $\widehat{\mathcal{I}}_{\mathrm{LC},\alpha}$ further reduces coverage error from $O(n^{-1})$ to $O(n^{-3/2})$.

A specific coverage error expression for the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_\alpha = (-\infty, q_\alpha(\hat{\theta})]$ is derived by Vidoni (1998). It is assumed that the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denotes n independent and identically distributed random variables drawn from a population with probability density function $f_X(x\,;\theta)$, where θ is the unknown parameter vector, and the predictand X is a random variable which is independent of $\mathcal{X}$ but drawn from the same population. Suppose $(\hat{\theta}, A)$ denotes a prediction-sufficient statistic, with A denoting an ancillary statistic. Then, as in Section 2.4, $P(X \in \widehat{\mathcal{I}}_\alpha)$ can be replaced by

$$\begin{aligned} P(X \in \widehat{\mathcal{I}}_\alpha \,|\, a) &= \int F_X\{q_\alpha(\vartheta)\,;\theta\} f_{\hat{\theta}}(\vartheta\,;\theta \,|\, a)\, d\vartheta \\ &= \int F_X\{q_\alpha(\vartheta)\,;\theta\} f^*_{\hat{\theta}}(\vartheta\,;\theta \,|\, a)\, d\vartheta + O(n^{-3/2}) \qquad (2.57) \\ &= \alpha + R\{q_\alpha(\theta)\,;\theta, a\} + O(n^{-3/2}), \qquad (2.58) \end{aligned}$$

where $f^*_{\hat{\theta}}(\cdot\,;\theta)$ is defined at (2.43),

$$\begin{aligned} R\{q_\alpha(\theta)\,;\theta,a\} = \tfrac{1}{2}\Big(\Big[\dot{f}_X\{q_\alpha(\theta)\,;\theta\} q_\alpha(\theta)_{/r} q_\alpha(\theta)_{/s} + f_X\{q_\alpha(\theta)\,;\theta\} q_\alpha(\theta)_{/rs}\Big]\check{\jmath}^{rs} \\ - f_X\{q_\alpha(\theta)\,;\theta\} q_\alpha(\theta)_{/r} \check{\imath}_{s;\,tu} \check{\jmath}^{rs} \check{\jmath}^{tu}\Big), \qquad (2.59) \end{aligned}$$

$\dot{f}_X(x\,;\theta) = \partial f_X(x\,;\theta)/\partial x$, $\check{\imath}_{R_1;\,R_2}$ is defined at (2.48) for index sets R_1 and R_2, $\check{\jmath}^{ij}$ denotes the (i,j)-th element from the inverse of the matrix $[-\check{\imath}_{ij;}]$, and $q_\alpha(\theta)_{/R}$ denotes the partial derivative of the α-th quantile of $F_X(x\,;\theta)$ with respect to the components of the parameter vector θ given by the index set R. One proceeds from (2.57) to (2.58) via an application of Laplace's method for integrals, as illustrated in Vidoni (1995) and referred to in Section 2.3.

Differentiating the identity $F_X\{q_\alpha(\theta);\theta\} = \alpha$ with respect to the r-th component of θ delivers the equation

$$f_X\{q_\alpha(\theta);\theta\}q_\alpha(\theta)_{/r} + F_X\{q_\alpha(\theta);\theta\}_{;\,r} = 0, \tag{2.60}$$

where

$$F_X(x;\theta)_{R_1;\,R_2} = F_X(x;\theta)_{/R_1;\,R_2}$$

represents the partial derivative of F_X with respect to the components of x and θ represented by the index sets R_1 and R_2. Furthermore, differentiating (2.60) with respect to the s-th component of θ provides, assuming the orders of differentiation can be interchanged, the equation

$$\begin{aligned}\dot{f}_X\{q_\alpha(\theta);\theta\}q_\alpha(\theta)_{/r}q_\alpha(\theta)_{/s} + f_X\{q_\alpha(\theta);\theta\}q_\alpha(\theta)_{/rs} \\ + f_X\{q_\alpha(\theta);\theta\}_{;\,r}q_\alpha(\theta)_{/s}[2] + F_X\{q_\alpha(\theta);\theta\}_{;\,rs} = 0,\end{aligned} \tag{2.61}$$

where

$$f_X(x;\theta)_{R_1;\,R_2} = f_X(x;\theta)_{/R_1;\,R_2}$$

represents the partial derivative of f_X with respect to the components of x and θ represented by the index sets R_1 and R_2. Using (2.60) and (2.61) it follows that (2.59) may be written in the equivalent form

$$\begin{aligned}R\{q_\alpha(\theta);\theta,a\} = -\tfrac{1}{2}\Big[F_X\{q_\alpha(\theta);\theta\}_{;\,rs}\check{j}^{rs} - F_X\{q_\alpha(\theta);\theta\}_{;\,r}\check{l}_{s,tu}\check{j}^{rs}\check{j}^{tu} \\ - 2F_X\{q_\alpha(\theta);\theta\}_{;\,r}f_X\{q_\alpha(\theta);\theta\}_{;\,s}\check{j}^{rs}/f\{q_\alpha(\theta);\theta\}\Big].\end{aligned} \tag{2.62}$$

Vidoni (1998) further shows that

$$P(X \in \widehat{\mathcal{I}}_\alpha|a) = \alpha + \tilde{R}\{q_\alpha(\theta);\theta,a\} + O(n^{-3/2}), \tag{2.63}$$

where

$$\begin{aligned}\tilde{R}\{q_\alpha(\theta);\theta,a\} = -\tfrac{1}{2}\Big[F_X\{q_\alpha(\theta);\theta\}_{;\,rs}i^{rs} - F_X\{q_\alpha(\theta);\theta\}_{;\,r}(i_{s,tu} + i_{s,t,u})i^{rs}i^{tu} \\ - 2F_X\{q_\alpha(\theta);\theta\}_{;\,r}f_X\{q_\alpha(\theta);\theta\}_{;\,s}i^{rs}/f_X\{q_\alpha(\theta);\theta\}\Big]\end{aligned} \tag{2.64}$$

is obtained by replacing $\check{j}^{tu}$ by i^{tu} and $\check{l}_{s,tu}$ by $i_{s,tu} + i_{s,t,u}$ in (2.62) (Barndorff-Nielsen and Cox, 1994, Section 5.5). Here

$$i_{s,tu} = E\{l_{/s;}\,(\theta;\hat{\theta},a)l_{/tu;}\,(\theta;\hat{\theta},a)\}, \quad i_{s,t,u} = E\{l_{/s;}\,(\theta;\hat{\theta},a)l_{/t;}\,(\theta;\hat{\theta},a)l_{/u;}\,(\theta;\hat{\theta},a)\},$$

and i^{ij} denotes the (i,j)-th element of the inverse of the matrix $[-E\{l_{/ij;}(\theta\,;\hat{\theta},a)\}]$.

In an attempt to achieve further reduction in coverage error, an additive-calibrated version of $\widehat{\mathcal{I}}_\alpha = (-\infty, q_\alpha(\hat{\theta})]$, given by

$$\widehat{\mathcal{I}}_{\mathrm{AC},\alpha} = (-\infty, q_\alpha(\hat{\theta}) + r(\alpha\,;\hat{\theta},a)],$$

is proposed by Vidoni (1998) to extirpate the $O(n^{-1})$ term, $\tilde{R}\{q_\alpha(\theta)\,;\theta,a\}$, from expansion (2.63). Under appropriate regularity conditions, and by direct analytic investigation, it was shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{AC},\alpha} \,|\, a) = \alpha + O(n^{-3/2})$$

whenever

$$r(\alpha\,;\tilde{\theta},a) = -\tilde{R}\{q_\alpha(\hat{\theta})\,;\hat{\theta},a\}/f_X\{q_\alpha(\hat{\theta})\,;\hat{\theta}\}. \tag{2.65}$$

Therefore, $\widehat{\mathcal{I}}_{\mathrm{AC},\alpha}$ further reduces conditional, and hence unconditional, coverage error from $O(n^{-1})$ to $O(n^{-3/2})$.

Assume that

$$F_X(x\,;\theta)_{;\,r} = \int_{-\infty}^{x} f_X(y\,;\theta)_{;\,r}\,dy \quad \text{and} \quad F_X(x\,;\theta)_{;\,rs} = \int_{-\infty}^{x} f_X(y\,;\theta)_{;\,rs}\,dy \tag{2.66}$$

and let $l_X(\theta\,;x) = \log f_X(x\,;\theta)$. Then, direct differentiation gives

$$f_X(x\,;\theta)_{;\,r} = l_X(\theta\,;x)_{r;}\, f_X(x\,;\theta) \tag{2.67}$$

and

$$f_X(x\,;\theta)_{;\,rs} = \{l_X(\theta\,;x)_{rs;} + l_X(\theta\,;x)_{r;}\, l_X(\theta\,;x)_{s;}\}\, f_X(x\,;\theta), \tag{2.68}$$

where $l_X(\theta\,;x)_{R_1;\,R_2} = l_X(\theta\,;x)_{/R_1;\,R_2}$ denotes the partial derivative of $l_X(\theta\,;x)$ with respect to the components of θ and x given by the index sets R_1 and R_2 respectively. When $r(\alpha\,;\hat{\theta},a)$ is given by (2.65), and employing (2.66) in (2.64) and further manipulating this expression using (2.60), (2.67), and (2.68), it can be shown that

$$q_\alpha(\hat{\theta}) + r(\alpha\,;\hat{\theta},a) = q_\alpha(\hat{\theta}) + q_\alpha(\hat{\theta})_{/r} l_X\{\hat{\theta}\,;q_\alpha(\hat{\theta})\}_{s;}\, \hat{i}^{rs} - f_X\{q_\alpha(\hat{\theta})\,;\hat{\theta}\}^{-1} \int_{-\infty}^{q_\alpha(\hat{\theta})} \tilde{S}(x\,;\hat{\theta},a) f_X(x\,;\hat{\theta})\,dx, \tag{2.69}$$

where $\hat{i}^{rs}$ is equivalent to i^{rs} except all occurrences of θ are replaced by $\hat{\theta}$, and

$$\tilde{S}(x\,;\theta,a) = -\tfrac{1}{2}\{l_X(\theta\,;x)_{rs;}\, i^{rs} + l_X(\theta\,;x)_{r;}\, l_X(\theta\,;x)_{s;}\, i^{rs} - l_X(\theta\,;x)_{r;}\, (i_{s,tu} + i_{s,t,u})i^{rs}i^{tu}\}.$$

From the explicit expression for $q_\alpha(\hat{\theta}) + r(\alpha\,;\hat{\theta},a)$ given by (2.69), consider the retrospective construction of a predictive density $\check{p}(x\,;\mathfrak{x})$, where $\hat{\theta} = \hat{\theta}(\mathfrak{x})$ and $a = a(\mathfrak{x})$ are explicitly dependent on the sample, such that

$$\int_{-\infty}^{q_\alpha(\hat{\theta})+r(\alpha\,;\hat{\theta},a)} \check{p}(x\,;\mathfrak{x})\,dx = \alpha,$$

$\check{p}(x\,;\mathfrak{x}) \geq 0$, and $\int \check{p}(x\,;\mathfrak{x})\,dx = 1$. To identify $\check{p}$, proceed by employing implicit differentiation and Taylor expansion (Barndorff-Nielsen and Cox, 1996, Section 6). It was shown by Vidoni (1998), under (2.66) and appropriate regularity conditions, that

$$\check{p}(x\,;\mathfrak{x}) = p_{\mathrm{EQI}}(x\,;\mathfrak{x}) + O(n^{-3/2}), \tag{2.70}$$

where $p_{\mathrm{EQI}}(x\,;\mathfrak{x})$ denotes the conditional predictive density for the predictand X and is defined by

$$p_{\mathrm{EQI}}(x\,;\mathfrak{x}) = f_X\{x\,;\check{\theta}(\alpha)\}\big[1 + \tilde{S}\{x\,;\check{\theta}(\alpha),a\}\big]$$

with $\check{\theta}(\alpha) = \hat{\theta} + [\hat{i}^{rs}]\nabla_\theta l_X\{\hat{\theta}\,;q_\alpha(\hat{\theta})\}$. See Vidoni (1998) for further alterations to $\tilde{S}\{x\,;\check{\theta}(\alpha),a\}$ which maintain the order of approximation given at (2.70). Note that $\int p_{\mathrm{EQI}}(x\,;\mathfrak{x})\,dx = 1$ and, in contrast to the quantile inferred predictive density $p_{\mathrm{QI}}(x\,;\mathfrak{x})$ (see (2.55)), $p_{\mathrm{EQI}}(x\,;\mathfrak{x})$ depends on the nominal level α.

The former analytic-calibrated one-sided prediction intervals, while effective at reducing coverage error, require the statistician to have a high level of technical expertise for implementation and also assume the predictand X is a random variable. When the predictand X is a random vector the statistician has two choices; either reformulate into a number of predictand random variables or consider a formulation of a prediction region which is uniquely determined. For the latter case, Hall, Peng, and Tajvidi (1999) investigate the coverage error of two nominal α-level prediction regions for the predictand X derived via profiling the approximate predictive likelihood and the estimative predictive density. A form of empirical calibration — the parametric bootstrap — is also considered which can be implemented with minimal technical fuss. Details for this approach are as follows.

Suppose the sample matrix $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denotes n independent and identically distributed random vectors from a population with probability density function $f(\cdot\,;\theta)$. Let the

predictand X denote a random vector of length l which is independent of $\mathcal{X}$ but drawn from a population which has $f_X(x;\theta)$ as probability density function. Under these assumptions, the approximate predictive likelihood for the predictand X, defined by (2.35) and proposed by Davison (1986), can be expressed as

$$p_{\mathrm{AP}}(x;\mathfrak{x}) = \frac{\det\{\hat{J}(\hat{\theta})\}^{1/2} f_X(x;\hat{\theta}_x)\Pi_{i=1}^n f(x_i;\hat{\theta}_x)}{\det\{\hat{J}_X(\hat{\theta}_x)\}^{1/2}\Pi_{i=1}^n f(x_i;\hat{\theta})},$$

with $p_{\mathrm{SAP}}(x;\mathfrak{x}) = p_{\mathrm{AP}}(x;\mathfrak{x})/\int p_{\mathrm{AP}}(y;\mathfrak{x})\,dy$ denoting the standardised version of the approximate predictive likelihood. Recall that $\hat{\theta} = \arg\max_\theta f_{\mathcal{X}}(\mathfrak{x};\theta) = \arg\max_\theta \Pi_{i=1}^n f(x_i;\theta)$, $\hat{\theta}_x = \arg\max_\theta f_{(\mathcal{X},X)}(\mathfrak{x},x;\theta) = \arg\max_\theta f_X(x;\theta)\Pi_{i=1}^n f(x_i;\theta)$, and, $\hat{J}(\theta)$ and $\hat{J}_X(\theta)$ denote the Hessian matrices, with respect to the parameter vector θ, of $-\sum_{i=1}^n \log f(x_i;\theta)$ and $-\log f_X(x;\theta) - \sum_{i=1}^n \log f(x_i;\theta)$, respectively.

Define the nominal α-level prediction regions for the predictand X, obtained by profiling the standardised approximate prediction likelihood $p_{\mathrm{SAP}}(x;\mathfrak{x})$ and the estimative predictive density $p_{\mathrm{E}}(x;\hat{\theta}) = f_X(x;\hat{\theta})$, defined at (2.40), by

$$\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha} = \{x : p_{\mathrm{SAP}}(x;\mathfrak{x}) > c_{\mathrm{SAP},\alpha}\} \quad \text{and} \quad \widehat{\mathcal{R}}_{\mathrm{E},\alpha} = \{x : p_{\mathrm{E}}(x;\mathfrak{x}) > c_{\mathrm{E},\alpha}\},$$

respectively. The constants $c_{\mathrm{SAP},\alpha}$ and $c_{\mathrm{E},\alpha}$ are selected such that

$$\int_{\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}} p_{\mathrm{SAP}}(x;\mathfrak{x})\,dx = \int_{\widehat{\mathcal{R}}_{\mathrm{E},\alpha}} p_{\mathrm{E}}(x;\mathfrak{x})\,dx = \alpha.$$

Under appropriate regularity conditions it was shown by Hall, Peng, and Tajvidi (1999) that

$$P(X \in \widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}) = \alpha + n^{-1}\pi_{\mathrm{SAP}} + O(n^{-2}) \quad \text{and} \quad P(X \in \widehat{\mathcal{R}}_{\mathrm{E},\alpha}) = \alpha + n^{-1}\pi_{\mathrm{E}} + O(n^{-2}), \quad (2.71)$$

where

$$\pi_{\mathrm{SAP}} = \int_{\mathcal{R}_\alpha} \left[\mathrm{tr}\psi(x) - \mathrm{tr}\{J_0^{-1} H_\theta f_X(x;\theta_0)\} - \frac{1}{2}\varphi(x) \right] dx + \alpha \int \varphi(x)\, dx,$$

$$\pi_{\mathrm{E}} = \frac{1}{2} \int_{\mathcal{R}_\alpha} \left[\nabla_\theta f(x;\theta_0)^\top J_0^{-1} \int \frac{\nabla_\theta f(y;\theta_0)\eta(y)}{f(y;\theta_0)}\, dy - \mathrm{tr}\{J_0^{-1} H_\theta f_X(x;\theta_0)\} \right] dx$$

$$+ \lim_{n\to\infty} nE\left[\left(\int_{\mathcal{R}_\alpha \cap \widehat{\mathcal{R}}_{\mathrm{E},\alpha}^c} - \int_{\widehat{\mathcal{R}}_{\mathrm{E},\alpha} \cap \mathcal{R}_\alpha^c} \right) (\hat\theta - \theta_0)_{(r)} \nabla_\theta f_X(x;\theta_0)^{(r)}\, dx \right],$$

$$\varphi(x) = \frac{\nabla_\theta f_X(x;\theta_0)^\top J_0^{-1} \nabla_\theta f_X(x;\theta_0)}{f_X(x;\theta_0)},$$

$$\psi(x) = \int J_0^{-1} \nabla_\theta f(y;\theta_0) \nabla_\theta f_X(x;\theta_0)^\top J_0^{-1} \left\{ \frac{H_\theta f(y;\theta_0)}{f(y;\theta_0)} - \frac{\nabla_\theta f(y;\theta_0) \nabla_\theta f(y;\theta_0)^\top}{f(y;\theta_0)^2} \right\} dy,$$

$$\eta(y) = \mathrm{tr}\{J_0^{-1} H_\theta f(y;\theta_0)\},$$

$$\hat J(\theta) = \sum_{i=1}^n \frac{\nabla_\theta f(X_i;\theta) \nabla_\theta f(X_i;\theta)^\top}{f(X_i;\theta)^2} - \sum_{i=1}^n \frac{H_\theta f(X_i;\theta)}{f(X_i;\theta)},$$

$$J_0 = \tfrac{1}{n} E\{\hat J(\theta_0)\},$$

$\mathrm{tr}A$ denotes the trace of the matrix A, $\mathcal{A}^c$ denotes the compliment in $\mathbb{R}^l$ of the set $\mathcal{A}$, $\mathcal{R}_\alpha = \{x : f_X(x;\theta_0) > c_\alpha\}$ with c_α selected such that $\int_{\mathcal{R}_\alpha} f_X(x;\theta_0)\, dx = \alpha$, and, $(\hat\theta - \theta_0)_{(r)}$ and $\nabla_\theta f_X(x;\theta_0)^{(r)}$ denote the r-th components of $\hat\theta - \theta_0$ and $\nabla_\theta f_X(x;\theta_0)$, respectively. Therefore, the coverage error associated with both nominal α-level prediction regions $\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}$ and $\widehat{\mathcal{R}}_{\mathrm{E},\alpha}$ is $O(n^{-1})$.

In an attempt to achieve further reduction in coverage error, level-calibrated versions of $\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}$ and $\widehat{\mathcal{R}}_{\mathrm{E},\alpha}$, given by

$$\widehat{\mathcal{R}}_{\mathrm{LC,SAP},\alpha} = \widehat{\mathcal{R}}_{\mathrm{SAP},\hat\beta_{\mathrm{SAP}}} \quad \text{and} \quad \widehat{\mathcal{R}}_{\mathrm{LC,E},\alpha} = \widehat{\mathcal{R}}_{\mathrm{E},\hat\beta_{\mathrm{E}}}, \tag{2.72}$$

respectively, are proposed by Hall, Peng, and Tajvidi (1999) to extirpate the $O(n^{-1})$ terms, $n^{-1}\pi_{\mathrm{SAP}}$ and $n^{-1}\pi_{\mathrm{E}}$, respectively, from the expansions at (2.71), where $\hat\beta_{\mathrm{SAP}}$ and $\hat\beta_{\mathrm{E}}$ are derived empirically using the parametric bootstrap as follows.

Having computed the maximum likelihood estimator $\hat\theta$ from the sample $\mathcal{X}$, draw a synthetic sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ which, conditional on $\mathcal{X}$, denotes n independent and identically distributed random vectors drawn from a population with probability density function $f(\cdot;\hat\theta)$. Use $\mathcal{X}^*$ to compute the associated maximum likelihood estimates $\hat\theta^* = \arg\max_\theta \Pi_{i=1}^n f(X_i^*;\theta)$ and $\hat\theta_x^* = \arg\max_\theta f_X(x;\theta)\Pi_{i=1}^n f(X_i^*;\theta)$. Set

$$\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}^* = \{x : p_{\mathrm{SAP}}(x;\mathcal{X}^*) > c_{\mathrm{SAP},\alpha}^*\} \quad \text{and} \quad \widehat{\mathcal{R}}_{\mathrm{E},\alpha}^* = \{x : p_{\mathrm{E}}(x;\mathcal{X}^*) > c_{\mathrm{E},\alpha}^*\},$$

where

$$p_{\mathrm{AP}}(x\,;\mathcal{X}^*) = \det\{\hat{J}(\hat{\theta}^*)\}^{1/2} f_X(x\,;\hat{\theta}_x^*)\Pi_{i=1}^n f(X_i^*\,;\hat{\theta}_x^*)/\big[\det\{\hat{J}_X(\hat{\theta}_x^*)\}^{1/2}\Pi_{i=1}^n f(X_i^*\,;\hat{\theta}^*)\big],$$

$$p_{\mathrm{SAP}}(x\,;\mathcal{X}^*) = \frac{p_{\mathrm{AP}}(x\,;\mathcal{X}^*)}{\int p_{\mathrm{AP}}(x\,;\mathcal{X}^*)\,dx},$$

$$p_{\mathrm{E}}(x\,;\mathcal{X}^*) = f_X(x\,;\hat{\theta}^*),$$

and nonnegative constants $c^*_{\mathrm{SAP},\alpha}$ and $c^*_{\mathrm{E},\alpha}$ satisfy

$$\int_{\widehat{\mathcal{R}}^*_{\mathrm{SAP},\alpha}} p_{\mathrm{SAP}}(x\,;\mathcal{X}^*)\,dx = \alpha \quad \text{and} \quad \int_{\widehat{\mathcal{R}}^*_{\mathrm{E},\alpha}} p_{\mathrm{E}}(x\,;\mathcal{X}^*)\,dx = \alpha,$$

respectively.

Draw a synthetic predictand X^* which, conditional on $\mathcal{X}$, is independent of the synthetic sample $\mathcal{X}^*$ and drawn from a population having $f_X(\cdot\,;\hat{\theta})$ as probability density function. Let

$$\hat{p}_{\mathrm{SAP}}(\beta) = P(X^* \in \widehat{\mathcal{R}}^*_{\mathrm{SAP},\beta}|\mathcal{X}) \quad \text{and} \quad \hat{p}_{\mathrm{E}}(\beta) = P(X^* \in \widehat{\mathcal{R}}^*_{\mathrm{E},\beta}|\mathcal{X})$$

for $\beta \in (0,1)$. Then sample quantities $\hat{\beta}_{\mathrm{SAP}}$ and $\hat{\beta}_{\mathrm{E}}$ of (2.72) are obtained as the calculated solutions of $\hat{p}_{\mathrm{SAP}}(\beta) = \alpha$ and $\hat{p}_{\mathrm{E}}(\beta) = \alpha$, respectively.

Under appropriate regularity conditions it was shown by Hall, Peng, and Tajvidi (1999) that

$$P(X \in \widehat{\mathcal{R}}_{\mathrm{LC,SAP},\alpha}) = \alpha + O(n^{-2}) \quad \text{and} \quad P(X \in \widehat{\mathcal{R}}_{\mathrm{LC,E},\alpha}) = \alpha + O(n^{-2}).$$

Therefore, the level-calibrated regions $\widehat{\mathcal{R}}_{\mathrm{LC,SAP},\alpha}$ and $\widehat{\mathcal{R}}_{\mathrm{LC,E},\alpha}$ of $\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}$ and $\widehat{\mathcal{R}}_{\mathrm{E},\alpha}$, respectively, further reduce coverage error, from $O(n^{-1})$ to $O(n^{-2})$.

Pragmatically, approximations to $\hat{p}_{\mathrm{SAP}}(\beta)$ and $\hat{p}_{\mathrm{E}}(\beta)$ are obtained by either direct numerical integration or repeated Monte Carlo simulation; approximations to $\hat{\beta}_{\mathrm{SAP},\alpha}$ and $\hat{\beta}_{\mathrm{E},\alpha}$ are then derived from the former approximations in conjunction with a numerical root finding technique.

In general, most computational labour in the construction of the prediction region $\widehat{\mathcal{R}}_{\mathrm{E},\alpha}$ occurs when calculating $\hat{\theta}$, which requires numerical maximisation techniques, and $c_{\mathrm{E},\alpha}$, which requires numerical root finding in conjunction with numerical integration that typically utilises quadrature, and is used to solve

$$\int_{\widehat{\mathcal{R}}_{\mathrm{E},\alpha}} p_{\mathrm{E}}(x\,;\mathfrak{x})\,dx = \alpha.$$

The construction of the prediction region $\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}$ proceeds in a similar way with, in general, additional computational labour required to calculate $\hat{\theta}_x$, which requires numerical maximisation

techniques, for either all abscissae points if using quadrature or integrand evaluations used in the numerical integration of

$$\int_{\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}} p_{\mathrm{SAP}}(x\,;\mathfrak{x})\,dx.$$

A major advantage of the prediction region $\widehat{\mathcal{R}}_{\mathrm{SAP},\alpha}$, or its calibrated version, would have to be perceived for it to be preferred to $\widehat{\mathcal{R}}_{\mathrm{E},\alpha}$ or its calibrated version, respectively, since these conformed prediction regions have commensurate coverage error.

When the predictand X is a random variable which is dependent on the sample $\mathcal{X}$, Beran (1990) considers

$$\widehat{\mathcal{R}}_\alpha = \{x\,:\,R(x,\mathcal{X}) \le F^{-1}_{R(X,\mathcal{X})|\mathcal{X}}(\alpha\,;\tilde{\theta}\,|\,\mathcal{X})\}$$

as a nominal α-level prediction region. Here, $R(X,\mathcal{X})$ is a real valued function of X and $\mathcal{X}$ selected by the statistician, $F_{R(X,\mathcal{X})|\mathcal{X}}\{R(x,\mathfrak{x})\,;\theta\,|\,\mathfrak{x}\}$ denotes the probability distribution function of $R(X,\mathcal{X})$ given the sample $\mathcal{X}=\mathfrak{x}$, $F^{-1}_{R(X,\mathcal{X})|\mathcal{X}}(\alpha\,;\theta\,|\,\mathfrak{x}) = \inf\{r\,:\,F_{R(X,\mathcal{X})|\mathcal{X}}(r\,;\theta\,|\,\mathfrak{x}) \ge \alpha\}$ denotes the inverse of $F_{R(X,\mathcal{X})|\mathcal{X}}$, and $\tilde{\theta}$ denotes a consistent estimator of the unknown parameter vector θ obtained using the sample $\mathcal{X}$ only.

Denote the probability distribution function of $F_{R(X,\mathcal{X})|\mathcal{X}}\{R(X,\mathcal{X})\,;\tilde{\theta},\mathcal{X}\}$ by

$$F_F(\cdot\,;\theta) = F_{F_{R(X,\mathcal{X})|\mathcal{X}}\{R(X,\mathcal{X})\,;\tilde{\theta}\,|\mathcal{X}\}}(\cdot\,;\theta),$$

and let $F_F^{-1}(\alpha\,;\theta) = \inf\{f\,:\,F_F(f\,;\theta) \ge \alpha\}$ denote the inverse of F_F. Under appropriate regularity conditions it was shown by Beran (1990) that

$$P(X \in \widehat{\mathcal{R}}_\alpha) = F_{F_{R(X,\mathcal{X})|\mathcal{X}}\{R(X,\mathcal{X})\,;\tilde{\theta}\,|\mathcal{X}\}}(\alpha\,;\theta) = \alpha + O(n^{-1}).$$

Therefore, the coverage error of $\widehat{\mathcal{R}}_\alpha$ is $O(n^{-1})$.

In an attempt to achieve further reduction in coverage error, a level-calibrated version of $\widehat{\mathcal{R}}_\alpha$, given by

$$\widehat{\mathcal{R}}_{\mathrm{LC},\alpha} = \widehat{R}_{F_F^{-1}(\alpha\,;\tilde{\theta})},$$

was proposed by Beran (1990). Under appropriate regularity conditions it was shown that

$$P(X \in \widehat{\mathcal{R}}_{\mathrm{LC},\alpha}) = \alpha + O(n^{-2}).$$

Therefore, the level-calibrated version $\widehat{\mathcal{R}}_{\mathrm{LC},\alpha}$ of $\widehat{\mathcal{R}}_\alpha$ further reduces coverage error from $O(n^{-1})$ to $O(n^{-2})$.

When probability distribution functions $F_{R(X,\mathcal{X})|\mathcal{X}}$ and F_F are difficult to ascertain, recourse can be had to Monte Carlo approximations (Beran, 1990, Section 2.3).

Chapter 3

Nonparametric Approaches

In this chapter, nonparametric approaches to the problem of constructing a prediction interval are defined and reviewed. Conceptual advantages for a nominal α-level prediction interval, for any $\alpha \in (0,1)$, are assessed via coverage error with other selection criteria being used when conformed prediction intervals have commensurate coverage errors.

This chapter is organised as follows. Section 3.1 considers a sample of independent and identically distributed random variables and treats two types of methods for constructing a nominal one-sided α-level prediction interval when the predictand is a random variable independent of the sample but drawn from the same population. The basis of one method is a Studentised statistic and has features in common with the construction of a confidence interval using the percentile-t method (Hall, 1992). The other method considers estimates of the α-th population quantile which are separately used as an upper end-point of a proposed nominal one-sided α-level prediction interval. Two further methods — percentile-t and accelerated bias-correction — are also considered for a predictand statistic of the predictand sample which is independent of the former sample but drawn from the same population.

Section 3.2 reviews regression and structural models in connection with construction of a nominal α-level prediction interval.

3.1 Independent and Identically Distributed Sample

Set $\alpha \in (0,1)$ and let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables from a population with F as probability distribution function. Before

directly considering the construction of prediction intervals, methodological philosophies from the construction of nominal α-level confidence intervals for a true population attribute $\theta = \theta(F)$ will be drawn up on.

Suppose the true population attribute $\theta = \theta(F)$ has value θ_0 and is estimated by $\hat{\theta} = \theta(F_n)$, where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{(-\infty,x]}(X_i)$ denotes the empirical distribution function. Define $\sigma^2 = \sigma^2(F)$ to be the asymptotic variance of $n^{1/2}\hat{\theta}$ and put $\hat{\sigma} = \sqrt{\sigma^2(F_n)}$.

Let the bootstrap sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ denote, conditional on the sample $\mathcal{X}$, n independent and identically distributed random variables drawn with replacement from $\mathcal{X}$, or equivalently, drawn from a population with F_n as probability distribution function. Then two nominal one-sided α-level confidence intervals for the true population attribute θ derived from the percentile and percentile-t methods are given by

$$\widehat{\mathcal{I}}_{\mathrm{CP},\alpha} = (-\infty, \hat{\theta} - n^{-1/2}\hat{\sigma}\hat{z}_{1-\alpha}] \quad \text{and} \quad \widehat{\mathcal{I}}_{\mathrm{CS},\alpha} = (-\infty, \hat{\theta} - n^{-1/2}\hat{\sigma}\hat{t}_{1-\alpha}], \tag{3.1}$$

respectively, where

$$\hat{z}_\alpha = \inf\Big\{z : P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \le z | \mathcal{X}\} \ge \alpha\Big\}, \tag{3.2}$$

$$\hat{t}_\alpha = \inf\Big\{t : P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \le t | \mathcal{X}\} \ge \alpha\Big\}, \tag{3.3}$$

$\hat{\theta}^* = \theta(F_n^*)$, $\hat{\sigma}^* = \sqrt{\sigma^2(F_n^*)}$, and $F_n^*(x) = n^{-1}\sum_{i=1}^{n} \mathbf{I}_{(-\infty,x]}(X_i^*)$. Confidence intervals $\widehat{\mathcal{I}}_{\mathrm{CP},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$ emanate from the theoretical one-sided α-level confidence intervals

$$\mathcal{I}_{\mathrm{CP},\alpha} = (-\infty, \hat{\theta} - n^{-1/2}\sigma z_{1-\alpha}] \quad \text{and} \quad \mathcal{I}_{\mathrm{CS},\alpha} = (-\infty, \hat{\theta} - n^{-1/2}\hat{\sigma} t_{1-\alpha}], \tag{3.4}$$

respectively, where z_α and t_α satisfy

$$P\{n^{1/2}(\hat{\theta} - \theta)/\sigma \le z_\alpha\} = \alpha \quad \text{and} \quad P\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \le t_\alpha\} = \alpha, \tag{3.5}$$

respectively.

When $n^{1/2}(\hat{\theta} - \theta_0)/\sigma$ and $n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma}$ satisfy the assumptions given by Hall (1992, Theorem 2.2) the following Edgeworth expansions can be developed:

$$P\{n^{1/2}(\hat{\theta} - \theta_0)/\sigma \le x\} = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + O(n^{-1}) \tag{3.6}$$

and

$$P\{n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma} \le x\} = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + O(n^{-1}), \tag{3.7}$$

where p_1 and q_1 are even polynomials of at most second degree, and, Φ and ϕ denote the standard normal distribution and density functions, respectively. Additionally, under the assumptions given by Hall (1992, Theorem 5.1), the following sample versions of the Edgeworth expansions given by (3.6) and (3.7) can be developed:

$$P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \leq x | \mathcal{X}\} = \Phi(x) + n^{-1/2}\hat{p}_1(x)\phi(x) + O_p(n^{-1}) \tag{3.8}$$

and

$$P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x | \mathcal{X}\} = \Phi(x) + n^{-1/2}\hat{q}_1(x)\phi(x) + O_p(n^{-1}), \tag{3.9}$$

where $\hat{p}_1$ and $\hat{q}_1$ are obtained from p_1 and q_1, respectively, by replacing population moments with sample moments. In particular, if the true population attribute is given by the population mean, $\theta = \int x\, dF(x)$, it can be shown that $p_1(x) = -\frac{1}{6}\gamma(x^2-1)$, $q_1(x) = \frac{1}{6}\gamma(2x^2+1)$, $\hat{p}_1(x) = -\frac{1}{6}\hat{\gamma}(x^2-1)$, and $\hat{q}_1(x) = \frac{1}{6}\hat{\gamma}(2x^2+1)$, where $\sigma^2 = \int (x-\theta)^2\, dF(x)$, $\hat{\theta} = n^{-1}\sum_{i=1}^n X_i$, $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \hat{\theta})^2$, $\gamma = E\big[\{(X_1 - \mu)/\sigma\}^3\big]$, and $\hat{\gamma} = \frac{1}{n}\sum_{i=1}^n \{(X_i - \hat{\theta})/\hat{\sigma}\}^3$.

By inverting (3.6) and (3.7) it follows that Cornish-Fisher expansions can be derived (Hall, 1992, Theorem 2.4) which stipulate that uniformly in $\epsilon < \alpha < 1 - \epsilon$ for each $0 < \epsilon < \frac{1}{2}$,

$$z_{1-\alpha} = w_{1-\alpha} + n^{-1/2}p_{11}(w_{1-\alpha}) + O(n^{-1}) \tag{3.10}$$

and

$$t_{1-\alpha} = w_{1-\alpha} + n^{-1/2}q_{11}(w_{1-\alpha}) + O(n^{-1}), \tag{3.11}$$

where $w_{1-\alpha}$ denotes the $(1-\alpha)$-th quantile of Φ, the standard normal distribution function,

$$p_{11}(x) = -p_1(x) \quad \text{and} \quad q_{11}(x) = -q_1(x).$$

By inverting (3.8) and (3.9) it follows that sample versions of the Cornish-Fisher expansions given by (3.10) and (3.11) can be derived (Hall, 1992, Theorem 5.2), which stipulate that uniformly in $\epsilon < \alpha < 1 - \epsilon$ for each $0 < \epsilon < \frac{1}{2}$,

$$\hat{z}_{1-\alpha} = w_{1-\alpha} + n^{-1/2}\hat{p}_{11}(w_{1-\alpha}) + O_p(n^{-1}) \tag{3.12}$$

and

$$\hat{t}_{1-\alpha} = w_{1-\alpha} + n^{-1/2}\hat{q}_{11}(w_{1-\alpha}) + O_p(n^{-1}), \tag{3.13}$$

where $\hat{p}_{11}$ and $\hat{q}_{11}$ are obtained from p_{11} and q_{11}, respectively, by replacing population moments with sample moments.

Under sufficiently stringent moment assumptions in conjunction with application of the Central Limit Theorem (Hall, 1992), it can be deduced that $\hat{p}_{11}(x) - p_{11}(x) = O_p(n^{-1/2})$ and $\hat{q}_{11}(x) - q_{11}(x) = O_p(n^{-1/2})$. Using (3.10) and (3.12),

$$\begin{aligned}\hat{z}_{1-\alpha} - z_{1-\alpha} &= n^{-1/2}\{\hat{q}_{11}(w_{1-\alpha}) - q_{11}(w_{1-\alpha})\} + O_p(n^{-1}) \\ &= n^{-1/2}O_p(n^{-1/2}) + O_p(n^{-1}) = O_p(n^{-1}),\end{aligned}$$

and similarly, using (3.11) and (3.13),

$$\hat{t}_{1-\alpha} - t_{1-\alpha} = O_p(n^{-1}). \tag{3.14}$$

Therefore the difference between the upper end-points for $\mathcal{I}_{\mathrm{CP},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CP},\alpha}$ satisfies

$$\begin{aligned}(\hat{\theta} - n^{-1/2}\hat{\sigma}\hat{z}_{1-\alpha}) - (\hat{\theta} - n^{-1/2}\sigma z_{1-\alpha}) &= n^{-1/2}\{(\sigma - \hat{\sigma})z_{1-\alpha} - \hat{\sigma}(\hat{z}_{1-\alpha} - z_{1-\alpha})\} \\ &= O_p(n^{-1})\end{aligned}$$

whenever $\hat{\sigma} - \sigma = O_p(n^{-1/2})$. In contrast, the difference between the upper end-points for $\mathcal{I}_{\mathrm{CS},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$ satisfies

$$(\hat{\theta} - n^{-1/2}\hat{\sigma}\hat{t}_{1-\alpha}) - (\hat{\theta} - n^{-1/2}\hat{\sigma}t_{1-\alpha}) = n^{-1/2}\hat{\sigma}(t_{1-\alpha} - \hat{t}_{1-\alpha}) = O_p(n^{-3/2}). \tag{3.15}$$

Under appropriate regularity conditions (Hall, 1992, Proposition 3.1),

$$P(\theta_0 \in \widehat{\mathcal{I}}_{\mathrm{CP},\alpha}) = \alpha + O(n^{-1/2}) \quad \text{and} \quad P(\theta_0 \in \widehat{\mathcal{I}}_{\mathrm{CS},\alpha}) = \alpha + O(n^{-1}). \tag{3.16}$$

Therefore, coverage error of $\widehat{\mathcal{I}}_{\mathrm{CP},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$ is $O(n^{-1/2})$ and $O(n^{-1})$, respectively.

Consider now a retrospective analysis of the confidence intervals for θ_0 defined at (3.1). These confidence intervals, given by $\widehat{\mathcal{I}}_{\mathrm{CP},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$, are based on estimating the $(1-\alpha)$-th quantiles, $\sigma z_{1-\alpha}$ and $t_{1-\alpha}$, of the non-Studentised statistic

$$n^{1/2}(\hat{\theta} - \theta_0) \tag{3.17}$$

and the Studentised statistic

$$\frac{n^{1/2}(\hat{\theta} - \theta_0)}{\hat{\sigma}}, \tag{3.18}$$

respectively, by the bootstrap estimators $\hat{\sigma}\hat{z}_{1-\alpha}$ and $\hat{t}_{1-\alpha}$, respectively, where $\hat{z}_\alpha$ and $\hat{t}_\alpha$ satisfy (3.2) and (3.3), respectively. Under appropriate regularity conditions, the Studentised statistic given by (3.18) is asymptotically pivotal in that it has the same limiting distribution for all values of unknowns. However, the non-Studentised statistic given by (3.17) is not asymptotically pivotal since it has a limiting distribution which depends on the unknown population attribute σ (Hall, 1992).

As enunciated by Hall (1992) it is the non-pivotal property of the non-Studentised statistic (3.17) which militates against $\widehat{\mathcal{I}}_{\mathrm{CP},\alpha}$ to deliver a coverage error of order only $n^{-1/2}$. In contrast, the pivotal property of the Studentised statistic (3.18) causes $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$ to deliver a coverage error of order n^{-1}. Therefore it is beneficial to construct a confidence interval for a true population attribute using a Studentised statistic when coverage error is to be reduced.

3.1.1 Studentised Method

Consider now the direct construction of a prediction interval predicated on the former methodological philosophies for confidence intervals. To this end, this subsection proposes a nominal α-level prediction interval which is constructed via a Studentised statistic, and investigates its properties. The bootstrap is used to estimate population quantities.

Let the predictand X denote a random variable which is independent of the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ but drawn from the same population with F as population distribution function. When constructing a prediction interval, the role of the predictand X is analogous to that of θ_0 for the confidence intervals discussed above. When $\theta = \int x\, dF(x)$ it follows that $\hat{\theta} = \bar{X} = n^{-1}\sum_{i=1}^n X_i$, and that a Studentised distribution can be proposed by replacing θ_0 by X and $\hat{\sigma}$ by an estimate of the asymptotic variance of $n^{1/2}(\bar{X} - X)$ in (3.18). Note that $\mathrm{Var}\{n^{1/2}(\bar{X} - X)\} = \sigma^2(1+n)$ is unbounded in n, where $\sigma^2 = \int (x-\theta)^2\, dF(x)$; however, $\sqrt{\mathrm{Var}\{n^{1/2}(\bar{X} - X)\}}/n^{1/2} = \sigma\sqrt{(1+n)/n} \to \sigma$ as $n \to \infty$. Therefore it is plausible to replace θ_0 by X and $n^{1/2}/\hat{\sigma}$ by the estimate of σ when $\theta = \int x\, dF(x)$, $\hat{\varsigma}$, in (3.18) to obtain the Studentised statistic

$$\frac{\bar{X} - X}{\hat{\varsigma}}, \tag{3.19}$$

where

$$\hat{\varsigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

The statistic at (3.19) was nominated by Bai and Olshen (1986) and Bai, Bickel, and Olshen (1990). Notice that (3.19) has the same form, up to a constant, as the pivotal transformation given by (2.6) for a normal population with unknown mean and variance.

Using the Studentised statistic (3.19) a theoretical one-sided α-level prediction interval for the predictand X is given by

$$\mathcal{I}_{\mathrm{S},\alpha} = (-\infty, \bar{X} - \hat{\varsigma} q_{1-\alpha}],$$

where q_α satisfies

$$P\{(\bar{X} - X)/\hat{\varsigma} \leq q_\alpha\} = \alpha.$$

Let the bootstrap predictand X^* denote a random variable which, conditional on the sample $\mathcal{X}$, is independent of the bootstrap sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ but drawn from the same population. Then a nominal one-sided α-level prediction interval can be derived from $\mathcal{I}_{\mathrm{S},\alpha}$ by estimating the population quantity $q_{1-\alpha}$ by the sample quantity $\hat{q}_{1-\alpha}$, and is given by

$$\widehat{\mathcal{I}}_{\mathrm{S},\alpha} = (-\infty, \bar{X} - \hat{\varsigma}\hat{q}_{1-\alpha}],$$

where

$$\hat{q}_\alpha = \inf\Big\{q : P\{(\bar{X}^* - X^*)/\hat{\varsigma}^* \leq q | \mathcal{X}\} \geq \alpha\Big\},$$

$\bar{X}^* = n^{-1}\sum_{i=1}^n X_i^*$, and $\hat{\varsigma}^* = \{n^{-1}\sum_{i=1}^n (X_i^* - \bar{X}^*)^2\}^{1/2}$.

Under appropriate regularity conditions, viz. $E(|X|^{10})$ is finite, the first derivative of F is strictly positive and the second derivative of F is bounded, it was shown by Bai and Olshen (1986) that

$$n^{1/2}(\hat{q}_\alpha - q_\alpha) \xrightarrow{\mathrm{d}} \tau Z,$$

where $\xrightarrow{\mathrm{d}}$ represents convergence in distribution, Z is a standard normal random variable and $\tau^2 = \mathrm{Var}\{\mathbf{I}_{(-\infty,-q_\alpha]}(X)/F'(-q_\alpha) + X - \frac{q_\alpha}{2}X^2\}$ with F' denoting the first derivative of F, the population distribution function. Hence, $\hat{q}_\alpha - q_\alpha = O_p(n^{-1/2})$. Therefore, the difference between the upper end-points for $\mathcal{I}_{\mathrm{S},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{S},\alpha}$ satisfies

$$(\bar{X} - \hat{\varsigma}\hat{q}_{1-\alpha}) - (\bar{X} - \hat{\varsigma} q_{1-\alpha}) = O_p(n^{-1/2}). \tag{3.20}$$

Again under appropriate regularity conditions, it was shown by Bai, Bickel, and Olshen (1990) that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{S},\alpha}) = \alpha + O(n^{-3/4+\gamma}) \tag{3.21}$$

for any $\gamma > 0$. Therefore, the coverage error of $\widehat{\mathcal{I}}_{\mathrm{S},\alpha}$ is $O(n^{-3/4+\gamma})$.

The probability order at (3.20) and the coverage error of $\widehat{\mathcal{I}}_{\mathrm{S},\alpha}$ are incommensurate with the analogous results for the former confidence interval $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$, given by (3.4), which has corresponding probability order and coverage error given by $O_p(n^{-3/2})$ and $O(n^{-1})$, respectively (see (3.14) and (3.16)). This disparity can be reconciled by investigating the limiting distributional results corresponding to the Studentised statistics used to construct $\widehat{\mathcal{I}}_{\mathrm{S},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{CS},\alpha}$. Under appropriate regularity conditions, $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ converges in distribution to a standard normal random variable (see (3.7)). However, while $(\bar{X} - X)/\hat{\varsigma}$ converges in distribution to a standard normal random variable whenever an underlying normal population is realised, more general population assumptions prevent this conclusion. Therefore an Edgeworth expansion for $(\bar{X} - X)/\hat{\varsigma}$, and the corresponding sample version for $(\bar{X}^* - X^*)/\hat{\varsigma}^*$, appear to be precluded. Hence, results (3.20) and (3.21) follow.

3.1.2 Quantile Estimation Methods

In this subsection, the upper end-point of nominal one-sided α-level prediction intervals are given by an estimate of the α-th population quantile. Three quantile estimates of the population quantile are considered. The first is the α-th quantile of the empirical distribution function; distributional properties of this estimate are investigated and rely on the Berry-Esseen Theorem. The second is the α-th quantile of a kernel distribution estimator, and the third is an estimate given by linear interpolation among two empirical distribution function quantiles. The coverage errors of those prediction intervals, determined via the empirical distribution function, follow in part from results of uniformly distributed order statistics.

This type of method is analogous, in some respects, to the percentile method used to construct a confidence interval (see (3.17) and following discussion) which is based on a non-Studentised statistic. Construction and properties of the proposed α-level prediction intervals proceed as follows.

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables drawn from a population with F as continuous probability distribution function. Let the predictand X denote a random variable independent of $\mathcal{X}$ but drawn from the same population.

Set $\alpha \in (0,1)$, define the quantile function for F by

$$F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\},$$

and let the empirical distribution function be denoted by $F_n(x) = n^{-1}\sum_{i=1}^{n} \mathbf{I}_{(-\infty,x]}(X_i)$.

A theoretical one-sided α-level prediction interval is given by $\mathcal{I}_{\mathrm{Q},\alpha} = (-\infty, F^{-1}(\alpha)]$ since, according to Reiss (1989, Criterion 1.2.3), $P(X \in \mathcal{I}_{\mathrm{Q},\alpha}) = F\{F^{-1}(\alpha)\} = \alpha$.

Let $\xi_\alpha = F^{-1}(\alpha)$. A nominal one-sided α-level prediction interval can be obtained by estimating the population quantile ξ_α using a modified form of the bootstrap principle (Hall, 1992, Section 1.2) and is given by

$$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q1},\alpha}],$$

where

$$\hat{\xi}_{\mathrm{Q1},\alpha} = \inf\{\xi : P(X^* \leq \xi \mid \mathcal{X}) \geq \alpha\}$$

and the bootstrap predictand X^* denotes, conditional on $\mathcal{X}$, a random variable drawn from a population with F_n as a probability distribution function. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$, with $X_{i:n}$ denoting the i-th order statistic of the sample $\mathcal{X}$. Then, by definition, $\hat{\xi}_{\mathrm{Q1},\alpha} = F_n^{-1}(\alpha) = X_{\lceil n\alpha \rceil:n}$. Therefore, the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ can be written as

$$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, X_{\lceil n\alpha \rceil:n}], \tag{3.22}$$

where the ceiling of $n\alpha$ is defined by $\lceil n\alpha \rceil = \min\{m : m \geq n\alpha\}$.

Under appropriate regularity conditions, the difference between the upper end-points for $\mathcal{I}_{\mathrm{Q},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ satisfies

$$X_{\lceil n\alpha \rceil:n} - \xi_\alpha = O\{(\log n/n)^{1/2}\} \quad \text{with probability one}$$

and

$$X_{\lceil n\alpha \rceil:n} - \xi_\alpha = O_p(n^{-1/2}). \tag{3.23}$$

A proof of the former result is provided by Serfling (1980, p. 96) and is achieved using Taylor expansion, Hoeffding's inequality, and the Borel-Cantelli lemma; see the proof of Theorem 4.1 in Subsection 4.3.1 for a generalisation of this useful result. Result (3.23) is a consequence of

$$X_{\lceil n\alpha \rceil:n} \sim AN[\xi_\alpha, \alpha(1-\alpha)/\{nf^2(\xi_\alpha)\}], \tag{3.24}$$

whenever F possesses a probability density function f which is continuous at ξ_α and positive in a neighbourhood of ξ_α. Using (3.24), $n^{1/2}(X_{\lceil n\alpha\rceil:n} - \xi_\alpha)$ converges in distribution to a random variable, Y say, which does not depend on n; therefore, $n^{1/2}(X_{\lceil n\alpha\rceil:n} - \xi_\alpha) = O_p(1)$ and (3.23) follows directly.

In order to prove (3.24) it suffices to show that, for every $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left[\frac{n^{1/2}\{F_n^{-1}(\alpha)-F^{-1}(\alpha)\}}{\sqrt{\alpha(1-\alpha)}/f(\xi_\alpha)} \le x\right] = \Phi(x),$$

where Φ is the probability distribution function of a standard normal random variable. Using properties intrinsic to any distribution function (Serfling, 1980, Subsection 1.1.4), and relabeling, it follows that

$$\begin{aligned}
P\left[\frac{n^{1/2}\left\{F_n^{-1}(\alpha)-F^{-1}(\alpha)\right\}}{\sqrt{\alpha(1-\alpha)}/f(\xi_\alpha)} \le x\right] &= P\left\{F_n^{-1}(\alpha) \le F^{-1}(\alpha) + \frac{\sqrt{\alpha(1-\alpha)}}{n^{1/2}f(\xi_\alpha)}x\right\} \\
&= P\left[\alpha \le F_n\left\{F^{-1}(\alpha) + \frac{\sqrt{\alpha(1-\alpha)}}{n^{1/2}f(\xi_\alpha)}x\right\}\right] \\
&= P(Y_{n,p} \ge n\alpha) \\
&= P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} \ge \frac{n\alpha-np}{\sqrt{np(1-p)}}\right\} \\
&= P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} \ge \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}, \qquad (3.25)
\end{aligned}$$

where $Y_{n,p}$ is a binomial random variable with parameters n and $p = F\left\{\xi_\alpha + \frac{\sqrt{\alpha(1-\alpha)}}{n^{1/2}f(\xi_\alpha)}x\right\}$, denoting the number of trials and the probability of success, respectively.

Directly from (3.25),

$$\begin{aligned}
\left|P\left[\frac{n^{1/2}\{F_n^{-1}(\alpha)-F^{-1}(\alpha)\}}{\sqrt{\alpha(1-\alpha)}/f(\xi_\alpha)} \le x\right] - \Phi(x)\right| &= \left|P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} \ge \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi(x)\right| \\
&= \left|1 - P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} < \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi(x)\right| \\
&= \left|P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} < \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi\left\{\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right. \\
&\qquad \left. + \Phi\left\{\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} + \Phi(x) - 1\right| \\
&= \left|P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} < \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi\left\{\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right. \\
&\qquad \left. + \Phi(x) - \Phi\left\{-\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right| \\
&\le \left|P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} < \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi\left\{\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right| \qquad (3.26) \\
&\qquad + \left|\Phi(x) - \Phi\left\{-\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right|. \qquad (3.27)
\end{aligned}$$

According to Chow and Teicher (1997, Chapter 9) the following Berry-Esseen Theorem holds:

Theorem 3.1 (Berry-Esseen) *If* $\{X_i, i \ge 1\}$ *are independent random variables,* $E(X_i) = 0$, $E(X_i^2) = \sigma_i^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 > 0$, $\Gamma_n^{2+\delta} = \sum_{i=1}^n E|X_i|^{2+\delta} \le \infty$, $n \ge 1$, *for some* $\delta \in (0,1]$ *and* $S_n = \sum_{i=1}^n X_i$, *there exists a universal constant* C_δ *such that*

$$\sup_{-\infty<x<\infty} \left|P(S_n/s_n < x) - \Phi(x)\right| \le C_\delta \left(\frac{\Gamma_n}{s_n}\right)^{2+\delta}.$$

Therefore,

$$\sup_{-\infty<x<\infty} \left|P\left\{\frac{Y_{n,p}-np}{\sqrt{np(1-p)}} < \frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\} - \Phi\left\{\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}}\right\}\right| \le C\frac{\gamma(p)}{n^{1/2}} \qquad (3.28)$$

where C is a constant and $\gamma(p) = \{(1-p)^2 + p^2\}/\{p(1-p)\}^{1/2}$.

The continuity of F, in conjunction with the existence of the derivative of F, denoted by f, in a neighbourhood of ξ_α, stipulates that

$$\begin{aligned}
-\frac{n^{1/2}(\alpha-p)}{\sqrt{p(1-p)}} &= \frac{n^{1/2}}{\sqrt{p(1-p)}}\left[F\left\{\xi_\alpha + \frac{\sqrt{\alpha(1-\alpha)}}{n^{1/2}f(\xi_\alpha)}x\right\} - F(\xi_\alpha)\right] \\
&\to \frac{x\frac{\sqrt{\alpha(1-\alpha)}}{f(\xi_\alpha)}}{\sqrt{\alpha(1-\alpha)}} f(\xi_\alpha) = x \qquad (3.29)
\end{aligned}$$

and

$$\gamma(p) \to \{(1-\alpha)^2 + \alpha^2\}/\{\alpha(1-\alpha)\}^{1/2} \tag{3.30}$$

as $n \to \infty$. From (3.30) and (3.29), in conjunction with (3.28) and the continuity of Φ, it follows that (3.26) and (3.27) both converge to zero as $n \to \infty$. This concludes the verification of (3.24).

Consider next the coverage error properties of the nominal α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$. Let $\mathcal{U} = (U_1, U_2, \ldots, U_n)^\top$ denote n independent and identically distributed random variables from a population with $F_{\mathrm{U}}(x) = x\mathbf{I}_{(0,1]}(x) + \mathbf{I}_{(1,\infty)}(x)$ as probability distribution function, and let $U_{1:n} < U_{2:n} < \cdots < U_{n:n}$, with $U_{i:n}$ denoting the i-th order statistic of $\mathcal{U}$. Denote by $\stackrel{d}{=}$ equality of random variables in distribution. Then, employing $F(X_{i:n}) \stackrel{d}{=} U_{i:n}$ (Reiss, 1989, Lemma 1.2.4) and $E(U_{i:n}) = \frac{i}{n+1}$, in conjunction with the tower property of conditional expectation (see (2.12)) with $\mathcal{G}_1 = \{\emptyset, \Omega\}$, $\mathcal{G}_2 = \sigma(\mathcal{X})$, observe that for any $\alpha \in (0,1)$,

$$\begin{aligned}
P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) &= P\{X \le F_n^{-1}(\alpha)\} \\
&= P\{X \le X_{\lceil n\alpha \rceil:n}\} \\
&= E\Big[E\{\mathbf{I}_{(-\infty, X_{\lceil n\alpha \rceil:n}]}(X) \,|\, \mathcal{X}\}\Big] \\
&= E\{F(X_{\lceil n\alpha \rceil:n})\} \\
&= E(U_{\lceil n\alpha \rceil:n}) \\
&= \frac{\lceil n\alpha \rceil}{n+1} \\
&= \frac{n\alpha + \alpha}{n+1} + O(n^{-1}) = \alpha + O(n^{-1}).
\end{aligned} \tag{3.31}$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ is given by $O(n^{-1})$, in contrast to $O(n^{-3/4+\gamma})$, for any $\gamma > 0$, for the former Studentised method at (3.21).

Nevertheless, when $\alpha = i/(n+1)$, where $1 \le i \le n$, it follows that $\lceil n\alpha \rceil = i$. Hence

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) = E(U_{i:n}) = \alpha.$$

Furthermore, for any $\alpha \in (0,1)$ and using arguments similar to those used to derive (3.31), it can be shown that

$$\limsup_{n\to\infty} n|P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) - \alpha| > 0.$$

In this way, a coverage error of $O(n^{-1})$ is the least possible for the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$.

Conceptually, if $\hat{F}(x)$ is an estimator of the population probability distribution function F, a one-sided prediction interval could be proposed by setting $\hat{F}^{-1}(\alpha)$ as its upper end-point. For example, the empirical distribution function F_n was employed when deriving $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, F_n^{-1}(\alpha)]$ at (3.22). Alternatively, a kernel distribution estimator for F could be employed and is determined from the kernel density estimator for f, the population probability density function, as follows.

Let $\hat{f}$ denote the kernel density estimator for f defined by

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{3.32}$$

where the kernel K is a probability density function, and the bandwidth $h \to 0$ as $n \to \infty$. Literature relating to kernel density estimation can be found in, for example, Silverman (1986) and Simonoff (1996).

The kernel distribution estimator, $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(y)\, dy$, of $F(x)$ the population probability distribution function, has α-th quantile denoted by

$$\hat{\tau}_\alpha = \hat{F}^{-1}(\alpha) = \inf\{\tau : \hat{F}(\tau) \geq \alpha\}.$$

Then a nominal one-sided α-level prediction interval is given by

$$\widehat{\mathcal{I}}_{\mathrm{K},\alpha} = (-\infty, \hat{\tau}_\alpha].$$

Under appropriate regularity conditions (Falk and Reiss, 1989), it can be shown that the difference between the upper end-points for $\mathcal{I}_{\mathrm{Q},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{K},\alpha}$ satisfies

$$\hat{\tau}_\alpha - \xi_\alpha = O_p(n^{-1/2} + h^2).$$

By Taylor expansion and (4.173) from the proof of Theorem 4.4 in Subsection 4.3.4, it can be shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{K},\alpha}) = \alpha + O(n^{-1} + h^2).$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{\mathrm{K},\alpha}$ is $O(n^{-1} + h^2)$, which is no better than the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ given by $O(n^{-1})$ at (3.31). Sufficient regularity conditions are that the population distribution function F has one bounded derivative in a neighbourhood of the α-th population quantile ξ_α, that $f(\xi_\alpha) > 0$, and that the kernel K of (3.32) is a nonnegative, symmetric, and compactly supported probability density function.

Instead of utilising one order statistic for the upper end-point of a nominal upper one-sided α-level prediction interval at (3.22), Beran and Hall (1993) propose the use of linear interpolation among two quantiles of the empirical distribution function, or equivalently, among two order statistics. Construction of this one-sided α-level prediction interval proceeds as follows.

Select m such that $m < \alpha(n+1) \leq m+1$ and put

$$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}],$$

where

$$\begin{aligned}\hat{\xi}_{\mathrm{Q2},\alpha} &= \{1+m-(n+1)\alpha\}F_n^{-1}(m/n) + \{(n+1)\alpha - m\}F_n^{-1}\{(m+1)/n\} \\ &= \{1+m-(n+1)\alpha\}X_{m:n} + \{(n+1)\alpha - m\}X_{m+1:n}.\end{aligned}$$

Under appropriate regularity conditions, Beran and Hall (1993) show that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}) = \alpha + O(n^{-2}). \tag{3.33}$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ is $O(n^{-2})$.

Beran and Hall (1993) prove (3.33) using Rényi's representation (Rényi, 1953) and an exposition is as follows. Let $\mathcal{Z} = (Z_1, Z_2, \ldots, Z_n)^\top$ denote n independent and identically distributed random variables drawn from a standard exponential population with $F_Z(z) = 1 - e^{-z}$ as the probability distribution function. Let $Z_{1:n} < Z_{2:n} < \cdots < Z_{n:n}$, with $Z_{i:n}$ denoting the i-th order statistic of $\mathcal{Z}$. To obtain this representation, and according to Reiss (1989, Theorem 1.2.5), the r-th order statistic of the sample $\mathcal{X}$ can be written as $X_{r:n} \stackrel{d}{=} F^{-1}(U_{r:n})$. From Reiss (1989, Lemma 1.2.1), the r-th order statistic of $\mathcal{Z}$ can be written as $Z_{r:n} \stackrel{d}{=} -\ln(U_{n-r+1:n})$. Combining these two results, it follows that the r-th order statistic of the sample $\mathcal{X}$ can identically be written as

$$X_{r:n} \stackrel{d}{=} F^{-1}\{\exp(-Z_{n-r+1:n})\}. \tag{3.34}$$

Set $Z_{0:n} = 0$. Then, according to Reiss (1989, Theorem 1.6.1), the collection of random variables

$$(n-s+1)(Z_{s:n} - Z_{s-1:n}),$$

where $1 \leq s \leq n$, are not only independent but each is distributed as a standard exponential random variable. Consequently, the r-th order statistic of $\mathcal{Z}$ can be written as

$$Z_{r:n} = \sum_{i=1}^{r}(Z_{i:n} - Z_{i-1:n}) \stackrel{d}{=} \sum_{i=1}^{r}\frac{\check{Z}_i}{n-i+1}, \tag{3.35}$$

say, where $\check{\mathcal{Z}} = (\check{Z}_1, \check{Z}_2, \ldots, \check{Z}_n)^\top$ is independent of $\mathcal{Z}$ and denotes n independent and identically distributed random variables drawn from a standard exponential population with $F_Z(z)$ as the probability distribution function.

Application of (3.35) to (3.34) allows the r-th order statistic of the sample $\mathcal{X}$ to be representable as

$$X_{r:n} \stackrel{d}{=} F^{-1}\left\{ \exp\left(- \sum_{i=1}^{n-r+1} \frac{\check{Z}_{n-r+1:n}}{i} \right) \right\}.$$

From this representation it can be shown that the sequence $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ has the Markov property, i.e. for $t \in \mathbb{R}$

$$P(X_{r:n} \leq t \,|\, X_{1:n} = x_{1:n}, \ldots, X_{r-1:n} = x_{r-1:n}) = P(X_{r:n} \leq t \,|\, X_{r-1:n} = x_{r-1:n}),$$

where $x_{1:n}$, $x_{2:n}$, $\ldots$, $x_{r-1:n}$, denote realisations of the order statistics $X_{1:n}$, $X_{2:n}$, $\ldots$, $X_{r-1:n}$.

3.1.3 Predictand Statistic

This subsection differs from the previous two subsections since it considers the construction of a one-sided prediction interval for a predictand statistic of the predictand sample, rather than for a predictand random variable drawn from the population. Two prediction intervals are proposed and derived from the percentile-t and the accelerated bias-correction (ABC) methods.

Confidence intervals based on the ABC method are transformation-respecting (Hall, 1992, Section 3.10) and have a reputation for dealing with difficult problems (e.g., constructing a confidence interval for a correlation coefficient). Related literature can be found in, for example, Efron (1987), Hall (1992), and Efron and Tibshirani (1993).

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ and the predictand sample $\mathcal{X}_\mathrm{P} = (X_{\mathrm{P},1}, X_{\mathrm{P},2}, \ldots, X_{\mathrm{P},m})^\top$ be independent and denote n and m, respectively, independent and identically distributed random variables drawn from a population with F as probability distribution function. Suppose the true population attribute is denoted by $\theta = \theta(F) = g(\mu)$, where g is a real-valued smooth function and $\mu = \int x\, dF(x)$. From $\mathcal{X}$ and $\mathcal{X}_\mathrm{P}$ two estimates of θ are given by $\hat{\theta}_n = \theta(F_n) = g(\bar{X}_n)$ and the predictand statistic $\hat{\theta}_{\mathrm{P},m} = \theta(F_{\mathrm{P},m}) = g(\bar{X}_{\mathrm{P},m})$, respectively, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\bar{X}_{\mathrm{P},m} = m^{-1}\sum_{i=1}^m X_{\mathrm{P},i}$, $F_n(x) = n^{-1}\sum_{i=1}^n \mathbf{I}_{(-\infty,x]}(X_i)$, and $F_{\mathrm{P},m}(x) = m^{-1}\sum_{i=1}^m \mathbf{I}_{(-\infty,x]}(X_{\mathrm{P},i})$. Denote the asymptotic variance of $(m^{-1} + n^{-1})^{-1/2}(\hat{\theta}_{\mathrm{P},m} - \hat{\theta}_n)$ by $\sigma^2 = \sigma^2(F) = h(\mu)$, for some

real-valued smooth function h, and estimate σ^2 by $\hat{\sigma}_n^2 = \sigma^2(F_n) = h(\bar{X}_n)$, which is based on the sample $\mathcal{X}$ only.

Consider now a prediction interval constructed using the percentile-t method. A theoretical one-sided α-level prediction interval for the predictand statistic $\hat{\theta}_{\mathrm{P},m}$ is of the form

$$\mathcal{I}_{\mathrm{ST},\alpha} = (-\infty, \hat{\theta}_n + (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n t_\alpha],$$

where t_α satisfies $P\{(m^{-1} + n^{-1})^{-1/2}(\hat{\theta}_{\mathrm{P},m} - \hat{\theta}_n)/\hat{\sigma}_n \leq t_\alpha\} = \alpha$. Conditional on the sample $\mathcal{X}$, let the bootstrap sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ and the bootstrap predictand sample $\mathcal{X}_{\mathrm{P}}^* = (X_{\mathrm{P},1}^*, X_{\mathrm{P},2}^*, \ldots, X_{\mathrm{P},m}^*)^\top$ be independent and denote n and m, respectively, independent and identically distributed random variables drawn with replacement from $\mathcal{X}$, or equivalently, drawn from a population with F_n as probability distribution function. Then a nominal one-sided α-level prediction interval for the predictand statistic $\hat{\theta}_{\mathrm{P},m}$ can be obtained by replacing t_α by the estimate $\hat{t}_\alpha$ and is given by

$$\widehat{\mathcal{I}}_{\mathrm{ST},\alpha} = (-\infty, \hat{\theta}_n + (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n \hat{t}_\alpha],$$

where

$$\hat{t}_\alpha = \inf\{t : P\{(m^{-1} + n^{-1})^{-1/2}(\hat{\theta}_{\mathrm{P},m}^* - \hat{\theta}_n^*)/\hat{\sigma}_n^* \leq t\} \geq \alpha\},$$

$\hat{\theta}_{\mathrm{P},m}^* = g(\bar{X}_{\mathrm{P},m}^*)$, $\hat{\theta}_n^* = g(\bar{X}_n^*)$, $\hat{\sigma}_n^* = \{h(\bar{X}_n^*)\}^{1/2}$, $\bar{X}_{\mathrm{P},m}^* = m^{-1}\sum_{i=1}^m X_{\mathrm{P},i}^*$, and $\bar{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$.

Under appropriate regularity conditions and by using similar techniques as for (3.15) and (3.16), it was shown by Mojirsheibani and Tibshirani (1996) that the difference between the upper end-points for $\mathcal{I}_{\mathrm{ST},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{ST},\alpha}$ satisfies

$$\begin{aligned}\{\hat{\theta}_n + (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n\hat{t}_\alpha\} - \{\hat{\theta}_n + (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n t_\alpha\} &= (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n(\hat{t}_\alpha - t_\alpha) \\ &= O_p\{\min(n,m)^{-3/2}\} \qquad (3.36)\end{aligned}$$

and that the coverage error of $\widehat{\mathcal{I}}_{\mathrm{ST},\alpha}$ is $O\{\min(m,n)^{-1}\}$, i.e.

$$P(\hat{\theta}_{\mathrm{P},m} \in \widehat{\mathcal{I}}_{\mathrm{ST},\alpha}) = \alpha + O\{\min(m,n)^{-1}\}. \qquad (3.37)$$

Consider now a prediction interval constructed using the ABC method. Let $\hat{H}(t) = P(\hat{\theta}_{\mathrm{P},m} \leq t \,|\, \mathcal{X})$, $z_0 = \Phi^{-1}\{P(\hat{\theta}_n^* < \hat{\theta}_n \,|\, \mathcal{X})\}$, $z_1 = \Phi^{-1}\{P(\hat{\theta}_{\mathrm{P},m}^* < \hat{\theta}_n \,|\, \mathcal{X})\}$, $a = \frac{1}{6}\big(\sum_{i=1}^n I_i^3\big)\big(\sum_{i=1}^n I_i^2\big)^{-3/2}$, and $b = a(n/m)^{1/2}$, where

$$I_i = \lim_{\epsilon \to 0} \frac{\theta\{(1-\epsilon)F_n + \epsilon\delta_i\} - \theta(F_n)}{\epsilon},$$

Φ denotes the standard normal distribution function, and $\delta_i(x) = \mathbf{I}_{\{x\}}(X_i)$. Denote by u_α and z_α the α-th quantile of $U = Z_1/Z_2$ and a standard normal random variable, respectively, where $Z_1 \sim N(1 - bz_1, b^2)$ and $Z_2 \sim N(1 - az_0, a^2)$. Then a nominal one-sided α-level prediction interval for the predictand statistic $\hat{\theta}_{\mathrm{P},m}$ is given by

$$\widehat{\mathcal{I}}_{\mathrm{ABC},\alpha} = (-\infty, \hat{w}_\alpha],$$

where

$$\hat{w}_\alpha = \begin{cases} \hat{H}^{-1}\Big\{\Phi\big(\frac{u_\alpha - 1}{b} + z_1\big)\Big\} & \text{if } a \neq 0, \\ \hat{H}^{-1}\Big[\Phi\big\{z_\alpha(1 + \frac{m}{n})^{1/2} + z_0(\frac{m}{n})^{1/2}\big\}\Big] & \text{otherwise.} \end{cases}$$

Under the same regularity conditions as (3.36) and (3.37), it was shown by Mojirsheibani and Tibshirani (1996) that the difference between the upper end-points for $\mathcal{I}_{\mathrm{ST},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{ABC},\alpha}$ satisfies

$$\hat{w}_\alpha - \{\hat{\theta}_n + (m^{-1} + n^{-1})^{1/2}\hat{\sigma}_n t_\alpha\} = O_p(n^{-3/2}) \tag{3.38}$$

and that the coverage error of $\widehat{\mathcal{I}}_{\mathrm{ABC},\alpha}$ is $O\{\min(m,n)^{-1}\}$, i.e.

$$P(\hat{\theta}_{\mathrm{P},m} \in \widehat{\mathcal{I}}_{\mathrm{ABC},\alpha}) = \alpha + O\{\min(m,n)^{-1}\}. \tag{3.39}$$

Since the prediction intervals $\widehat{\mathcal{I}}_{\mathrm{ST},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{ABC},\alpha}$ have commensurate probability orders at (3.36) and (3.38), in conjunction with commensurate coverage errors at (3.37) and (3.39), it is advisable to consider their other intrinsic properties. For example, a prediction interval $\widehat{\mathcal{I}}_{\mathrm{T},\alpha} = (-\infty, \hat{q}_\alpha(\mathcal{X})]$ constructed using method T, say, is range-preserving if $\hat{q}_\alpha(\mathcal{X})$ is contained in the support of F, the population probability distribution function. Additionally, $\widehat{\mathcal{I}}_{\mathrm{T},\alpha}$ is transformation-respecting for the predictand statistic $\hat{\theta}_{\mathrm{P},m}$ if given a monotonic function g (an increasing function, say) the prediction interval for $g(\hat{\theta}_{\mathrm{P},m})$ constructed using method T is given by $(\infty, g\{\hat{q}_\alpha(\mathcal{X})\}]$. Mojirsheibani and Tibshirani (1996) argue against the prediction interval $\widehat{\mathcal{I}}_{\mathrm{SP},\alpha}$, since it is not in general range-preserving or transformation-respecting. However, these properties are intrinsic to the prediction interval $\widehat{\mathcal{I}}_{\mathrm{ABC},\alpha}$.

3.2 Regression and Structural Models

This section generalises the population assumptions considered in Section 3.1. Regression and structural models will be reviewed in connection with construction of a nominal α-level prediction

interval. In both cases the bootstrap principle (Hall, 1992, Section 1.2) will be employed but is disguised and hampered by the problem of initial parameter estimation. A nominal one-sided α-level prediction interval is proposed for a structural model which is expected to have satisfactory coverage error.

Regression Model. Consider the linear regression model,

$$\mathcal{X} = Z\beta + \epsilon,$$

where the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denotes n independent and identically distributed random variables (observations), $Z = (z_1, z_2, \ldots, z_n)^\top$ denotes a fixed and nonstochastic $n \times k$ matrix with z_i a $k \times 1$ vector, β denotes a $k \times 1$ vector of unknown parameters, and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top$ denotes n independent and identically distributed random variables drawn from a population with probability distribution function F which has bounded continuous derivative $F' = f$, $f(w) > 0$ for all $w \in \mathbb{R}$, $E(\epsilon_i) = \int w\, dF(w) = 0$, $E(\epsilon_i^2) = \int w^2\, dF(w) = \sigma^2$, and $E(\epsilon_i^4) = \int w^4\, dF(w) < \infty$.

The least-squares estimator of β, $\hat{\beta}$, is obtained as the argument which minimises

$$S(\beta) = (\mathcal{X} - Z\beta)^\top (\mathcal{X} - Z\beta),$$

i.e. $\hat{\beta} = \arg\min_\beta S(\beta)$. Suppose that the $n \times 1$ vector $\mathbf{1} = (1, 1, \ldots, 1)^\top$ lies in the column span of Z, Z is of full rank, and that the eigenvalues of $Z^\top Z$ increase with n. Then the least-squares estimator of β is given by

$$\hat{\beta} = (Z^\top Z)^{-1} Z^\top \mathcal{X}, \tag{3.40}$$

and the residuals are denoted by $\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_n)^\top = \mathcal{X} - Z\hat{\beta}$.

Consider now the problem of constructing a prediction interval, based on the sample $\mathcal{X}$ of observations and Z the response matrix, for the predictand $X = z_{n+1}^\top \beta + \epsilon_{n+1}$, where z_{n+1} is a nonstochastic $k \times 1$ vector and ϵ_{n+1} denotes a random variable independent of ϵ but drawn from the same population with F as continuous distribution function. If F were known, then $G(w) = P(X - z_{n+1}^\top \hat{\beta} \leq w)$ would also be known. Therefore, a theoretical two-sided α-level prediction interval for X would be given by

$$\mathcal{I}_\alpha = [z_{n+1}^\top \hat{\beta} + G^{-1}\{(1-\alpha)/2\}, z_{n+1}^\top \hat{\beta} + G^{-1}\{(1+\alpha)/2\}],$$

where $G^{-1}(\alpha) = \inf\{w : G(w) \geq \alpha\}$.

When F is unknown, the bootstrap principle stipulates that the probability distribution function $G(w) = P(X - z_{n+1}^\top \hat{\beta} \le w)$ be replaced by an estimator $\hat{G}(w) = P\{X^* - z_{n+1}^\top \hat{\beta}^* \le w \mid (\mathcal{X}, Z)\}$, say, conditional on the sample $\mathcal{X}$ of observations and the response matrix Z, where the bootstrap predictand $X^* = z_{n+1}\hat{\beta} + \epsilon_{n+1}^*$, the bootstrap sample of observations $\mathcal{X}^* = Z\hat{\beta} + \epsilon^*$, the bootstrap least-squares parameter estimate $\hat{\beta}^* = (Z^\top Z)^{-1} Z^\top \mathcal{X}^*$, the bootstrap residuals $\epsilon^* = (\epsilon_1^*, \epsilon_2^*, \ldots, \epsilon_n^*)^\top$ denote, conditional on $(\mathcal{X}, Z)$, n independent and identically distributed random variables drawn from a population with

$$F_{\text{adj},n}(w) = n^{-1} \sum_{i=1}^{n} \mathbf{I}_{(-\infty, w]}(\hat{r}_i), \tag{3.41}$$

as probability distribution function, and, conditional on $(\mathcal{X}, Z)$, ϵ_{n+1}^* is independent of ϵ^* but drawn from the same population with $F_{\text{adj},n}$ as population distribution function. Here Stine (1985) selects the adjusted residuals $\hat{r}_i = \hat{\epsilon}_i/(1 - h_i)^{1/2}$ with $h_i = z_i^\top (Z^\top Z)^{-1} z_i$. The adjusted residuals $\hat{r}_i$ have zero mean and variance σ^2 and so do not have to be centered (Freedman, 1981) and exhibit first and second moments identical to the underlying population. A nominal two-sided α-level prediction interval for the predictand X is

$$\widehat{\mathcal{I}}_\alpha = [z_{n+1}^\top \hat{\beta} + \hat{G}^{-1}\{(1-\alpha)/2\}, z_{n+1}^\top \hat{\beta} + \hat{G}^{-1}\{(1+\alpha)/2\}],$$

where $\hat{G}^{-1}(\alpha) = \inf\{w : \hat{G}(w) \ge \alpha\}$, and it was shown by Stine (1985) that

$$P(X \in \widehat{\mathcal{I}}_\alpha) = \alpha + o(1).$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_\alpha$ converges to zero as $n \to \infty$. No explicit order of coverage error has been determined.

Structural Model. Consider a strictly stationary autoregressive process $(X_m\,; -\infty < m < \infty)$ of order $k \ge 1$ which satisfies

$$\varphi(B) X_m = \epsilon_m,$$

where B denotes the backward shift operator, i.e. $B\,X_m = X_{m-1}$, $\varphi(w) = 1 + \sum_{i=1}^{k} \theta_i w^i$, $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^\top$, $\varphi(w) \ne 0$ for all $w \in \mathbb{C}$ such that $|w| \le 1$, and $(\epsilon_m\,; -\infty < m < \infty)$ denotes a sequence of independent and identically distributed random variables drawn from a population with probability distribution function F such that $E(\epsilon_m) = \int w\, dF(w) = 0$ and $E(\epsilon_m^2) = \int w^2\, dF(w) =$

σ^2. The former restriction on $\varphi(w)$ engenders the representation

$$X_m = \sum_{i=0}^{\infty} \psi_i \epsilon_{m-i},$$

where $\sum_{i=0}^{\infty} \psi_i w^i = 1/\varphi(w)$ and $\sum_{i=0}^{\infty} |\psi_i| \leq \infty$ (Brockwell and Davis, 1987, Theorem 3.1.1).

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n sequential observations of the autoregressive process which will be used to construct $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)^\top$, the estimator of the parameter vector θ, and $\hat{F}$, the estimator of F. Various candidates for $\hat{\theta}$ are given in Fuller (1976) and Brockwell and Davis (1987, Chapter 8). When $\hat{\theta}$ denotes the least-squares estimate of θ, i.e.

$$\hat{\theta} = \arg\min_\theta S(\theta),$$

where $S(\theta) = (n-k)^{-1} \sum_{i=k+1}^{n} \{\varphi(B) X_i\}^2$, Thombs and Schucany (1990) propose

$$\hat{F}(w) = (n-k)^{-1} \sum_{i=k+1}^{n} \mathbf{I}_{(-\infty, w]}(\hat{\epsilon}_i)$$

as an estimator of F, where $\hat{\epsilon}_i = \hat{\epsilon}_i' - (n-k)^{-1} \sum_{j=k+1}^{n} \hat{\epsilon}_j'$, $\hat{\epsilon}_i' = \hat{\varphi}(B) X_i$, and $\hat{\varphi}(w) = 1 + \sum_{i=1}^{k} \hat{\theta}_i w^i$.

Consider now the problem of constructing a prediction interval, based on the sequential sample $\mathcal{X}$, for the predictand $X = X_{n+p}$, where the integer $p \geq 1$. A nominal one-sided α-level prediction interval for the predictand X would be of the form $\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{q}_\alpha(\mathcal{X})]$ where $\hat{q}_\alpha$ depends on the sample sequence $\mathcal{X}$ and α only. Suppose X_{P} is the minimum mean squared error predictor of X, i.e. the linear combination $X_{\mathrm{P}} = \sum_{i=1}^{n} \lambda_{n,i} X_{n+1-i}$ best approximates X in the sense that $E|X - X_{\mathrm{P}}|^2$ is minimised (Brockwell and Davis, 1987, Example 2.3.3). For an autoregressive process of order k and by using the Projection theorem (Brockwell and Davis, 1987, Theorem 2.3.1) it can be shown that $\lambda_{n,i} = 0$ whenever the integer $i > k$ (Brockwell and Davis, 1987, Example 5.3.1) and that in general the vector $\lambda = (\lambda_{n,1}, \lambda_{n,2}, \ldots, \lambda_{n,k})^\top$ depends on θ. Set $\hat{X}_{\mathrm{P}} = \sum_{i=1}^{n} \hat{\lambda}_{n,i} X_{n+1-i}$ where $\hat{\lambda} = (\hat{\lambda}_{n,1}, \hat{\lambda}_{n,2}, \ldots, \hat{\lambda}_{n,k})^\top$ denotes the vector λ with θ replaced by the estimate $\hat{\theta}$. Then Kabaila (1993) proposes that $\hat{q}_\alpha(\mathcal{X}) = \hat{X}_{\mathrm{P}} + \hat{c}_\alpha(\mathcal{X})$, where $\hat{c}_\alpha(\mathcal{X})$ depends on the sample sequence $\mathcal{X}$ and α only.

When $G(w) = P(X \leq w)$, the probability distribution function of X, is known and continuous, a theoretical one-sided α-level prediction interval for the predictand X is given by $\mathcal{I}_\alpha = (-\infty, G^{-1}(\alpha)]$. When G is unknown the bootstrap principle stipulates that it be replaced by an

estimator $\hat{G}(w) = P(X^* \leq w \,|\, \mathcal{X})$, where $X^* = X^*_{n+p}$,

$$X_i^* = \begin{cases} X_i & \text{if } 1 \leq i \leq k, \\ -\sum_{j=1}^{k} \hat{\theta}_j X^*_{i-j} + \epsilon_i^* & \text{otherwise,} \end{cases}$$

and $\epsilon^* = (\epsilon^*_{k+1}, \epsilon^*_{k+2}, \ldots, \epsilon^*_{n+p})$ denotes, conditional on $\mathcal{X}$, $n+p-k$ independent and identically distributed random variables drawn from a population with $\hat{F}$ as probability distribution function. Then a nominal one-sided α-level prediction interval is given by

$$\hat{\mathcal{I}}_\alpha = (-\infty, \hat{G}^{-1}(\alpha)],$$

where $\hat{G}^{-1}(\alpha) = \inf\{w : \hat{G}(w) \geq \alpha\}$.

If predictive inference is taken conditional on $(X_{n-k+1}, X_{n-k+2}, \ldots, X_n)$ it follows that the conditional distribution function $G(w) = P\{X \leq w \,|\, (X_{n-k+1}, X_{n-k+2}, \ldots, X_n)\}$ would be replaced by $\hat{G}(w) = P\{X^* \leq w \,|\, (X^*_{n-k+1}, X^*_{n-k+2}, \ldots, X^*_n) = (X_{n-k+1}, X_{n-k+2}, \ldots, X_n), \mathcal{X}\}$, where $X^* = X^*_{n+p}$, and as implicitly proposed by Kabaila (1993),

$$X_i^* = \begin{cases} X_i & \text{if } n-k+1 \leq i \leq n, \\ -\sum_{j=1}^{k} \hat{\theta}_j X^*_{i-j} + \epsilon_i^* & \text{if } i > n, \\ (\epsilon^*_{i+k} - X^*_{i+k} - \sum_{j=1}^{k-1} \hat{\theta}_j X^*_{i+k-j})/\hat{\theta}_k & \text{if } n-k \geq i \geq 1, \end{cases}$$

and $\epsilon^* = (\epsilon^*_{k+1}, \epsilon^*_{k+2}, \ldots, \epsilon^*_{n+p})$ denotes, conditional on $\mathcal{X}$, $n+p-k$ independent and identically distributed random variables drawn from a population with $\hat{F}$ as probability distribution function. When F or $\hat{F}$ is differentiable a Monte-Carlo estimator for G or $\hat{G}$, respectively, is proposed by Kabaila (1993).

Chapter 4

Nonparametric Prediction Intervals

The purpose of this chapter is to further investigate estimators of the α-th population quantile which are used as the upper end-point of proposed nominal one-sided α-level prediction intervals to achieve reduced coverage error. In particular, estimators are derived from interpolation among empirical distribution function quantiles. Under the assumptions considered in this chapter, empirical distribution function quantiles are equivalent to order statistics. Therefore, in addition to the estimates obtained from one order statistic and from linear interpolation among two order statistics (Chapter 3, Subsection 3.1.2), the interpolation among three or four order statistics will be considered in an attempt to reduce coverage error. Two types of calibration, the jackknife and the smoothed bootstrap, will be proposed for constructing nominal one-sided α-level prediction intervals.

Recall that two methods for constructing one-sided prediction intervals have already been discussed (Chapter 3, Subsections 3.1.1 and 3.1.2). Given the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ which denotes n independent and identically distributed random variables drawn from a population with F as probability distribution function and $\alpha \in (0,1)$, the first method involved the inversion of the Studentised statistic

$$g(X, \mathcal{X}) = \frac{\bar{X} - X}{\hat{\varsigma}}$$

with respect to the predictand X, which is a random variable independent of the sample $\mathcal{X}$ but drawn from the same population, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\hat{\varsigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. The

constructed nominal one-sided α-level prediction interval is then given by

$$\widehat{\mathcal{I}}_{\mathrm{S},\alpha} = (-\infty, \bar{X} - \hat{\varsigma}\hat{q}_{1-\alpha}],$$

where $\hat{q}_\alpha$ satisfies

$$\hat{q}_\alpha = \inf\{q : P\{(\bar{X}^* - X^*)/\hat{\varsigma}^* \leq q | \mathcal{X}\} \geq \alpha\}.$$

Here $\bar{X}^*$ and $\hat{\varsigma}^*$ denote the bootstrap versions of $\bar{X}$ and $\hat{\varsigma}$, respectively, calculated using the bootstrap sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ which denotes, conditional on $\mathcal{X}$, n independent and identically distributed random variables drawn from a population with $F_n(x) = n^{-1}\sum_{i=1}^n \mathbf{I}_{(-\infty,x]}(X_i)$ as probability distribution function, and the bootstrap predictand X^* denotes, conditional on $\mathcal{X}$, a random variable which is independent of $\mathcal{X}^*$ but drawn from the same population. Furthermore, as elucidated in Subsection 3.1.1 of Chapter 3, under appropriate regularity conditions it can be shown that, for any $\gamma > 0$,

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{S},\alpha}) = \alpha + O(n^{-3/4+\gamma}). \tag{4.1}$$

The second method considered estimates of the α-th population quantile $\xi_\alpha = F^{-1}(\alpha)$, constructed from the sample $\mathcal{X}$ and α only, which are used as an upper end-point for nominal one-sided α-level prediction intervals. For example, the α-th population quantile ξ_α, which satisfies $P(X \leq \xi_\alpha) = \alpha$ whenever F is continuous, can be estimated by the α-th bootstrap quantile $\hat{\xi}_{\mathrm{Q1},\alpha}$ which satisfies

$$\hat{\xi}_{\mathrm{Q1},\alpha} = \inf\{\xi : P(X^* \leq \xi \,|\, \mathcal{X}) \geq \alpha\}.$$

The constructed nominal one-sided α-level prediction interval is

$$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q1},\alpha}],$$

and, since $\hat{\xi}_{\mathrm{Q1},\alpha}$ can also be written as

$$\hat{\xi}_{\mathrm{Q1},\alpha} = F_n^{-1}(\alpha) = X_{\lceil n\alpha \rceil:n},$$

it can be shown (Chapter 3, Subsection 3.1.2) that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) = \alpha + O(n^{-1}). \tag{4.2}$$

Note that the quantile estimation and former Studentised methods are analogous to the percentile and percentile-t methods, respectively, for confidence interval construction (Section 3.1). However, while the coverage error of a one-sided confidence interval is of order n^{-1} or $n^{-1/2}$ when constructed via the percentile-t or percentile methods, respectively, the coverage error of a one-sided prediction interval is of order $n^{-3/4+\gamma}$, for any $\gamma > 0$, or n^{-1} when constructed via the Studentised or quantile estimation methods at (4.1) and (4.2). This disparity, in conjunction with additional results in Subsection 3.1.2, demonstrates the ease with which coverage error can be reduced by considering interpolation among order statistics as estimators of the α-th population quantile.

An outline of this chapter is as follows. Section 4.1 initially summarises the coverage error properties of nominal one-sided α-level prediction intervals constructed from quantile estimators based on one order statistic and the interpolation among two order statistics. Then, the coverage error properties of nominal one-sided α-level prediction intervals constructed from quantile estimators based on interpolation among three and four order statistics are investigated. Section 4.2 proposes two types of calibration, the jackknife and the smoothed bootstrap, for a nominal one-sided α-level prediction interval, in an attempt to further reduce coverage error. Section 4.3 considers proofs of referenced technical results.

4.1 Quantile Estimation via Interpolation among Order Statistics

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables drawn from a population with F as probability distribution function. The predictand X denotes a random variable which is independent of the sample $\mathcal{X}$ but drawn from the same population. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$, with $X_{i:n}$ denoting the i-th order statistic of $\mathcal{X}$, and set $\alpha \in (0,1)$.

Denote the α-th population quantile of F by ξ_α, i.e.

$$\xi_\alpha = F^{-1}(\alpha) = \inf\{\xi : F(\xi) \geq \alpha\}.$$

Furthermore, assume that the α-th population quantile ξ_α is uniquely defined. A necessary condition for ξ_α to be uniquely defined is that F be strictly increasing. If F is continuous and strictly increasing then F^{-1} is the standard inverse of F. When F is continuous it follows that

$F(\xi_\alpha) = P(X \leq \xi_\alpha) = \alpha$ (Reiss, 1989, Criterion 1.2.3). Therefore, $\mathcal{I}_{Q,\alpha} = (-\infty, \xi_\alpha]$ is a theoretical one-sided α-level prediction interval.

Given the integer $1 \leq q \leq n$, suppose that $1 \leq n_1 < n_2 < \cdots < n_q \leq n$. Then the α-th population quantile ξ_α may, in general, be estimated by

$$\hat{\xi}_{q,\alpha} = \sum_{i=1}^{q} a_i X_{n_i:n}, \tag{4.3}$$

where $a_i = O(1)$ for $i = 1, 2, \ldots, q$, $\sum_{i=1}^{q} a_i = 1 + o(1)$, and each a_i, n_i depends on α and n only. Set $\widehat{\mathcal{I}}_{q,\alpha} = (-\infty, \hat{\xi}_{q,\alpha}]$; from the results in Subsections 3.1.1 and 3.1.2 of Chapter 3, it is plausible to expect that for fixed q it would be possible to specify a_i and n_i such that

$$P(X \in \widehat{\mathcal{I}}_{q,\alpha}) = \alpha + O(n^{-q}). \tag{4.4}$$

In this section, specialisations of (4.3) are considered with $q = 1, 2, 3, 4$. Successive reductions in coverage error, from $O(n^{-1})$ to $O(n^{-2})$ and $O(n^{-3})$, are achieved when considering the quantile bootstrap estimator, followed by two and three point interpolation among order statistics, respectively. However, four point interpolation among order statistics cannot engender a prediction interval which has $O(n^{-4})$ as coverage error. Concise forms of these results are given in Theorem 4.1 and Theorem 4.2.

4.1.1 One Order Statistic

As previously discussed in Subsection 3.1.2 of Chapter 3, the α-th population quantile ξ_α may be estimated by the α-th bootstrap quantile defined by

$$\hat{\xi}_{Q1,\alpha} = F_n^{-1}(\alpha) = X_{m:n}, \tag{4.5}$$

where $m = \lceil n\alpha \rceil$. Note that (4.5) is a specialisation of (4.3), obtained when $q = 1$, $n_1 = m$, and $a_1 = 1$. The corresponding nominal one-sided α-level prediction interval is

$$\widehat{\mathcal{I}}_{Q1,\alpha} = (-\infty, \hat{\xi}_{Q1,\alpha}], \tag{4.6}$$

and it can be shown (Chapter 3, Subsection 3.1.2) that $\widehat{\mathcal{I}}_{Q1,\alpha}$ has coverage error of order n^{-1}, i.e.

$$P(X \in \widehat{\mathcal{I}}_{Q1,\alpha}) = \alpha + O(n^{-1}). \tag{4.7}$$

Additionally, assuming that the α-th population quantile ξ_α is uniquely defined and the population distribution function F is continuous, it can be shown that not only does (4.7) hold but

$$\limsup_{n\to\infty} n\left|P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) - \alpha\right| > 0.$$

Thus, the remainder $O(n^{-1})$, given at (4.7), is the least possible.

Suppose that the α-th population quantile ξ_α is estimated by the m-th order statistic of the sample $\mathcal{X}$ for some as yet unspecified m. In this case it can be shown that the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, X_{m:n}]$ is of order n^{-1}, i.e. $P(X \in \widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}) = \alpha + O(n^{-1})$, whenever $m = (n+1)\alpha + O(1)$. Verification of this result is similar to the proof of (4.7) (see Chapter 3, Subsection 3.1.2).

If $m = \lceil n\alpha \rceil$ this method has the desirable property that $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ can be constructed whenever the sample size $n \geq 1$, for any $\alpha \in (0,1)$.

4.1.2 Interpolation among Two Order Statistics

Consider the specialisation of (4.3) obtained when $q = 2$, $n_1 = m$, $n_2 = m+1$, $a_1 = a$, and $a_2 = b$. Denote this preliminary estimator of the α-th population quantile ξ_α by

$$\hat{\xi}_{2,\alpha} = aX_{m:n} + bX_{m+1:n} \tag{4.8}$$

and set

$$\widehat{\mathcal{I}}_{2,\alpha} = (-\infty, \hat{\xi}_{2,\alpha}].$$

Following the prescription at (4.4), it is possible to specify m, a, and b such that the coverage error of $\widehat{\mathcal{I}}_{2,\alpha}$ is of order n^{-2}.

From Theorem 4.1 in Section 4.3 of this chapter, it can be shown that

$$P(X \in \widehat{\mathcal{I}}_{2,\alpha}) = \frac{m+b}{n+1} + O(n^{-2})$$

whenever $a+b = 1$. Sufficient regularity conditions are that the population distribution function F has two bounded derivatives in a neighbourhood of the α-th population quantile ξ_α, and $F'(\xi_\alpha) > 0$. Therefore, the condition required for $\widehat{\mathcal{I}}_{2,\alpha}$ to enjoy coverage error of order n^{-2} is for

$$(m+b)/(n+1) = \alpha,$$

or equivalently,

$$a = m + 1 - (n+1)\alpha \qquad \text{and} \qquad b = (n+1)\alpha - m$$

whenever $a + b = 1$. Since the preliminary estimator of the α-th population quantile at (4.8) involves two sequential order statistics, and $0 \leq a,\ b \leq 1$, it follows that

$$m = \lfloor (n+1)\alpha \rfloor \qquad \text{and} \qquad 1 \leq m \leq n-1. \tag{4.9}$$

Define the estimator of the α-th population quantile by

$$\hat{\xi}_{\mathrm{Q2},\alpha} = \{m + 1 - (n+1)\alpha\} X_{m:n} + \{(n+1)\alpha - m\} X_{m+1:n} \tag{4.10}$$

and set the nominal one-sided α-level prediction interval

$$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}], \tag{4.11}$$

where m satisfies (4.9). From the former discussion, it follows that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}) = \alpha + O(n^{-2}).$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ is given by $O(n^{-2})$.

Note that for fixed $\alpha \in (0,1)$ the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, defined by (4.11), is available whenever m satisfies (4.9). For example, if $\alpha = 0.9,\ 0.95$ it is required that $n \geq 10,\ 20$, respectively, for $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ to exist. When, for fixed α, m does not satisfy (4.9), recourse can be had to the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ defined by (4.6), since it is defined for all $n \geq 1$ and $\alpha \in (0,1)$.

The function `quantile`, defined in the statistical software package S-PLUS[1], estimates the α-th population quantile ξ_α using $\hat{\xi}_{\mathrm{Q2},\alpha}$, given by (4.10), whenever $1 \leq m \leq n-1$ or using $\hat{\xi}_{\mathrm{Q1},\alpha}$, given by (4.5), otherwise.

4.1.3 Interpolation among Three Order Statistics

Consider the specialisation of (4.3) obtained when $q = 3$, $n_1 = m-1$, $n_2 = m$, $n_3 = m+1$, $a_1 = a$, $a_2 = b$, and $a_3 = c$. Denote this preliminary estimator of the α-th population quantile ξ_α by

$$\hat{\xi}_{3,\alpha} = aX_{m-1:n} + bX_{m:n} + cX_{m+1:n}, \tag{4.12}$$

and set

$$\widehat{\mathcal{I}}_{3,\alpha} = (-\infty, \hat{\xi}_{3,\alpha}].$$

[1] S-PLUS is a trademark of StatSci.

Following the prescription at (4.4), it is possible to specify m, a, b, and c such that the coverage error of $\widehat{\mathcal{I}}_{3,\alpha}$ is of order n^{-3}.

From Theorem 4.1 in Section 4.3 of this chapter, it can be shown that

$$P(X \in \widehat{\mathcal{I}}_{3,\alpha}) = \frac{m+c-a}{n+1} + \frac{A}{n^2}\{(a-c)^2 + a^2 + c^2 - 2(a+c)\} + O(n^{-3})$$

whenever $a+b+c=1$, where $A \in \mathbb{R}$. Sufficient regularity conditions are that the population distribution function F has three bounded derivatives in a neighbourhood of the α-th population quantile ξ_α and that $F'(\xi_\alpha) > 0$. Therefore the conditions required of $\widehat{\mathcal{I}}_{3,\alpha}$ to obtain a coverage error of order n^{-3}, are:

$$\frac{m+c-a}{n+1} = \alpha \tag{4.13}$$

and

$$(a-c)^2 + a^2 + c^2 - 2(a+c) = 0. \tag{4.14}$$

Using (4.13) and (4.14) to solve for a, b, c it follows that

$$a = \tfrac{1}{2}\{2 - C \pm (4-3C^2)^{1/2}\},\ b = -\{1 \pm (4-3C^2)^{1/2}\},\ c = \tfrac{1}{2}\{2 + C \pm (4-3C^2)^{1/2}\} \tag{4.15}$$

whenever $a+b+c=1$, where $C = (n+1)\alpha - m$ and the $+$ or $-$ sign is to be used throughout in the definitions of a, b, c. However, since a, b, $c \in \mathbb{R}$, it follows that $4 - 3C^2 \geq 0$, or equivalently,

$$|m - (n+1)\alpha| \leq \tfrac{2}{\sqrt{3}}. \tag{4.16}$$

It is not possible for the triplet (a,b,c) to satisfy the conditions $a+b+c=1$ and $0 \leq a$, b, $c \leq 1$ through selection of the quadruplet (m,a,b,c) except when $m = (n+1)\alpha$, in which case set $a=0$, $b=1$, $c=0$. To see why, note that $0 \leq b < 1$ when the $-$ sign is selected throughout in (4.15) whenever $-1 \geq C > 0$ or $0 < C \leq 1$ only. However, $0 < a \leq 1$ whenever $-1 \geq C > 0$ and $0 < c \leq 1$ whenever $0 < C \leq 1$. When $|C| \leq 1$ it is possible to show that $b \in [0,1]$ while a, $c \in [1 - 2/\sqrt{3}, 1]$.

When m and a, b, c satisfy (4.16) and (4.15), respectively, denote an estimator of the α-th population quantile ξ_α by

$$\hat{\xi}_{\mathrm{Q3},\alpha} = aX_{m-1:n} + bX_{m:n} + cX_{m+1:n} \tag{4.17}$$

and set the nominal one-sided α-level prediction interval

$$\widehat{\mathcal{I}}_{Q3,\alpha} = (-\infty, \hat{\xi}_{Q3,\alpha}]. \tag{4.18}$$

Then, from the former discussion, it follows that

$$P(X \in \widehat{\mathcal{I}}_{Q3,\alpha}) = \alpha + O(n^{-3}).$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{Q3,\alpha}$ is given by $O(n^{-3})$.

Note that for each m satisfying (4.16) there are two possible choices for the triple (a, b, c) which correspond to selecting the + or the − sign, respectively, throughout in (4.15). Additionally, as long as $2/(n+1) \leq \alpha \leq (n-2)/(n+1)$, there are three values of m which satisfy (4.16). Therefore, implicit in (4.17) is a certain amount of ambiguity; the statistician's dilemma is to select one quadruplet (m, a, b, c), which is consistent with (4.15) and (4.16), out of, at most, the six possible.

Only partial answers to this dilemma are available. If it is supposed that the sampling distribution is supported on the positive half-line then it is desired to select an m which satisfies (4.16) and coefficients a, b, c, which are given by (4.15) with the + or the − sign being selected throughout, such that, with probability one,

$$\hat{\xi}_{Q3,\alpha} > 0. \tag{4.19}$$

Let $C = (n+1)\alpha - m$. Then, for each $2/(n+1) \leq \alpha \leq (n-2)/(n+1)$, there exists a unique value of m such that

$$0 \leq C < 1. \tag{4.20}$$

For this value of m select

$$a = \tfrac{1}{2}\{2 - C - (4 - 3C^2)^{1/2}\},\ b = -\{1 - (4 - 3C^2)^{1/2}\},\ c = \tfrac{1}{2}\{2 + C - (4 - 3C^2)^{1/2}\}, \tag{4.21}$$

i.e. select the − sign throughout in (4.15). Then, for this quadruplet (m, a, b, c) and after rearranging, observe that

$$\hat{\xi}_{Q3,\alpha} = -a(X_{m:n} - X_{m-1:n}) + X_{m:n} + c(X_{m+1:n} - X_{m:n}). \tag{4.22}$$

Note that $-a$ and c are both nonnegative when $0 \leq C < 1$, and (4.22) is a nonnegative linear combination of nonnegative quantities. Only for this specific choice of m, $-a$, and c does (4.22) satisfy (4.19).

Suppose, alternatively, that the sampling distribution has support on the negative half-line. Then it is desired to select an m which satisfies (4.16), and coefficients a, b, c, which are given by (4.15) with the + or the − sign being selected throughout, such that, with probability one,

$$\hat{\xi}_{\mathrm{Q3},\alpha} < 0. \tag{4.23}$$

Then for each $2/(n+1) \leq \alpha \leq (n-2)/(n+1)$ there exists a unique value of m such that

$$-1 < C \leq 0.$$

For this value of m, select (a, b, c) as at (4.21). Then, for this quadruplet (m, a, b, c) and after rearranging, observe that

$$\hat{\xi}_{\mathrm{Q3},\alpha} = a(X_{m-1:n} - X_{m:n}) + X_{m:n} - c(X_{m:n} - X_{m+1:n}). \tag{4.24}$$

Note that a and $-c$ are both nonnegative when $-1 < C \leq 0$ and (4.24) is a nonpositive linear combination of nonpositive quantities. Only for this specific choice of m, a, and $-c$ does (4.24) satisfy (4.23).

When there is no underlying constraint on the sample, such as nonnegativity or nonpositivity, the statistician may consider selecting the quadruplet (m, a, b, c), such selection being consistent with (4.15) and (4.16), for the estimator (4.17) such that the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$ has smallest or largest upper end-point. Or one may consider a transformation of the sample which is supported on the positive half-line or negative half-line and use the estimator given by (4.22) or (4.24), respectively. Alternatively, assuming that the α-th population quantile estimator is consistent with the underlying support of the population, for given sample size n a specific quadruplet (m, a, b, c), such selection being consistent with (4.15) and (4.16), may be selected on the grounds of a favourable simulation study.

Note that for fixed $\alpha \in (0, 1)$ the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, defined by (4.18), is available whenever m satisfies (4.16) and

$$2 \leq m \leq n - 1. \tag{4.25}$$

The constraint at (4.25) is essential since the preliminary estimator of the α-th population quantile at (4.12) involves three sequential order statistics. For example, if $\alpha = 0.9, 0.95$ it is required that $n \geq 8$, 16, respectively, for $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$ to exist. When a, b, c are given by (4.21), $0 \leq C < 1$, and

$\alpha = 0.9, 0.95$ the prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$ exists and has a nonnegative end-point when the sample size $n \geq 10, 20$, respectively. When, for fixed α and triplet (a, b, c), m can not be selected to satisfy (4.16) and (4.25), recourse can be had to the nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ or $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ given by (4.6) or (4.11), respectively.

4.1.4 Interpolation among Four Order Statistics

Consider the specialisation of (4.3) obtained when $q = 4$, $n_1 = m$, $n_2 = m + 1$, $n_3 = m + 2$, $n_4 = m + 3$, $a_1 = a$, $a_2 = b$, $a_3 = c$, and $a_4 = d$. Denote this preliminary estimator of the α-th population quantile ξ_α by

$$\hat{\xi}_{4,\alpha} = aX_{m:n} + bX_{m+1:n} + cX_{m+2:n} + dX_{m+3:n}, \tag{4.26}$$

and set

$$\widehat{\mathcal{I}}_{4,\alpha} = (-\infty, \hat{\xi}_{4,\alpha}].$$

Following the prescription at (4.4) it is attempted to specify m, a, b, c, and d such that the coverage error of $\widehat{\mathcal{I}}_{4,\alpha}$ is of order n^{-4}.

From Theorem 4.2 in Section 4.3 of this chapter, it can be shown that

$$\begin{aligned} P(X \in \widehat{\mathcal{I}}_{4,\alpha}) = {} & \frac{m + b + 2c + 3d}{n+1} \\ & + \frac{\gamma_1(n)}{2}\{2(b^2 - b) + 6(c^2 - c) + 12(d^2 - d) + 6bc + 8bd + 16cd\} \\ & + \frac{1}{6}\Big[\gamma_2(n)\{6(b - b^2) + 24(c - c^2) + 120(d - d^2) - 20bc - 30bd - 70cd\} \\ & + \gamma_3(n)\{6(b^3 - b) + 24(c^3 - c) + 120(d^3 - d) + 24b^2c + 30b^2d \\ & \qquad + 36c^2b + 90c^2d + 120cd^2 + 60bd^2 + 90bcd\}\Big] + O(n^{-4}) \end{aligned}$$

whenever $a + b + c + d = 1 + o(1)$, where $\gamma_1(n) = O(n^{-2})$, $\gamma_2(n) = O(n^{-3})$, and $\gamma_3(n) = O(n^{-3})$. Sufficient regularity conditions are that the population distribution function F has four bounded derivatives in a neighbourhood of the α-th population quantile ξ_α and that $F'(\xi_\alpha) > 0$. Therefore

the conditions required of $\widehat{\mathcal{I}}_{4,\alpha}$, to obtain a coverage error of order n^{-4}, are:

$$b + 2c + 3d = C, \tag{4.27}$$

$$2(b^2 - b) + 6(c^2 - c) + 12(d^2 - d) + 6bc + 8bd + 16cd = 0, \tag{4.28}$$

$$6(b - b^2) + 24(c - c^2) + 120(d - d^2) - 20bc - 30bd - 70cd = 0, \tag{4.29}$$

$$6(b^3 - b) + 24(c^3 - c) + 120(d^3 - d) + 24b^2c + 30b^2d + 36c^2b + 90c^2d + 120cd^2 + 60bd^2 + 90bcd = 0, \tag{4.30}$$

where $C = (n+1)\alpha - m$.

It is stated in Theorem 4.2 that for any fixed $\alpha \in (0,1)$ there exists no simultaneous solution to (4.27), (4.28), (4.29), and (4.30) for the quintuplet (m, a, b, c, d). Hence, $\widehat{\mathcal{I}}_{4,\alpha}$ can never have coverage error of order n^{-4}. This result essentially follows from the fact that m is an integer.

As long as the quintuplet (m, a, b, c, d) is selected such that (4.27) and (4.28) are satisfied, the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{4,\alpha}$ will have coverage error of order n^{-3}. This is most likely accomplished via an additional constraint on the quadruplet (a, b, c, d).

To illustrate this point by an example, suppose that an estimator of the α-th population quantile ξ_α is given by (4.26) and that

$$c - d = 0. \tag{4.31}$$

Let $\hat{\xi}_{\mathrm{Q4C},\alpha} = aX_{m:n} + bX_{m+1:n} + cX_{m+2:n} + cX_{m+3:n}$ and set $\widehat{\mathcal{I}}_{\mathrm{Q4C},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q4C},\alpha}]$, where m satisfies

$$C \in [\tfrac{26}{19} - \tfrac{14}{19}\sqrt{5}, \tfrac{26}{19} + \tfrac{14}{19}\sqrt{5}], \tag{4.32}$$

and

$$a = \tfrac{1}{14}\big(26 - 5C \pm 3\sqrt{-19C^2 + 52C + 16}\big),$$
$$b = \tfrac{-1}{14}\big(20 + C \pm 5\sqrt{-19C^2 + 52C + 16}\big),$$
$$c = d = \tfrac{1}{14}\big(4 + 3C \pm \sqrt{-19C^2 + 52C + 16}\big),$$

where the $+$ or $-$ sign to be used throughout in the definitions of the triplet (a, b, c). From the former discussion and Theorem 4.2, it follows that $\hat{\xi}_{\mathrm{Q4C},\alpha}$ satisfies (4.31) and that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q4C},\alpha}) = \alpha + O(n^{-3}).$$

Therefore, the coverage error of $\widehat{\mathcal{I}}_{\text{Q4C},\alpha}$ is given by $O(n^{-3})$.

Note that for fixed $\alpha \in (0,1)$ the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\text{Q4C},\alpha}$ is available whenever m satisfies (4.32) and

$$1 \leq m \leq n-3. \tag{4.33}$$

The constraint at (4.33) is essential since the preliminary estimator of the α-th population quantile at (4.26) involves four sequential order statistics. For example, if $\alpha = 0.9, 0.95$ it is required that $n \geq 9$, 19, respectively, for $\widehat{\mathcal{I}}_{\text{Q4C},\alpha}$ to exist. When, for fixed α and triplet (a,b,c), m cannot be selected to satisfy (4.32) and (4.33), recourse can be had to the nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{\text{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\text{Q2},\alpha}$ or $\widehat{\mathcal{I}}_{\text{Q3},\alpha}$ given by (4.6), (4.11) or (4.18), respectively.

4.2 Prediction Interval Calibration

Let $\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{\xi}_\alpha]$ denote a generic nominal one-sided α-level prediction interval constructed using the sample $\mathcal{X}$ and α only, where $\hat{\xi}_\alpha$ denotes an estimator of the α-th population quantile. Then, in principle it is possible to select a tuning multiplicative, m, a tuning additive, a, and a tuning level, β, such that

$$P\big(X \in (-\infty, m\hat{\xi}_\alpha]\big) = \alpha,$$
$$P\big(X \in (-\infty, \hat{\xi}_\alpha + a]\big) = \alpha,$$
$$P\big(X \in (-\infty, \hat{\xi}_\beta]\big) = \alpha.$$

The nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_\alpha$ is multiplicative, additive or level calibrated whenever the population tuners m, a, and β are replaced by estimates $\hat{m}$, $\hat{a}$, and $\hat{\beta}$, respectively, producing multiplicative, additive or level calibrated prediction intervals given by $\widehat{\mathcal{I}}_{\text{MC},\alpha} = (-\infty, \hat{m}\hat{\xi}_\alpha]$, $\widehat{\mathcal{I}}_{\text{AC},\alpha} = (-\infty, \hat{\xi}_\alpha + \hat{a}]$ or $\widehat{\mathcal{I}}_{\text{LC},\alpha} = (-\infty, \hat{\xi}_{\hat{\beta}}]$, respectively. The purpose of multiplicative, additive or level calibration is to produce a coverage error,

$$P(X \in \widehat{\mathcal{I}}_{\text{MC},\alpha}) - \alpha, \quad P(X \in \widehat{\mathcal{I}}_{\text{AC},\alpha}) - \alpha \quad \text{or} \quad P(X \in \widehat{\mathcal{I}}_{\text{LC},\alpha}) - \alpha,$$

respectively, which is of smaller order than $P(X \in \widehat{\mathcal{I}}_\alpha) - \alpha$, else the respective calibration is ineffective.

When considering a true population attribute the iterated bootstrap is used for additive or level calibration of bootstrap confidence intervals and is successful at reducing coverage error and also

improving level error accuracy in bootstrap hypothesis testing (Hall, 1992; Efron and Tibshirani, 1993; Shao and Tu, 1995; Davison and Hinkley, 1997). Because of this, Hall (1992) states, "it would be hard " to exclude the use of bootstrap calibration when pursuing "utilitarian techniques".

This section considers two types of estimation procedures used for calibration. The first procedure, delineated in Subsection 4.2.1, considers the level calibration of a nominal α-level prediction interval where the estimator of the level tuner is obtained using the jackknife. The investigation of the former procedure is motivated by its beneficial computational advantage. The second procedure, delineated in Subsection 4.2.2, considers the additive calibration of a nominal α-level prediction interval where the estimate of the additive tuner is obtained using the smoothed bootstrap. A sample constructed using the smoothed bootstrap has no data ties; this property is not shared by the regular bootstrap. Reference material for the jackknife and the smoothed bootstrap can be found in, for example, Efron (1979, 1981, 1982), Efron and Gong (1983), Silverman (1986), Silverman and Young (1987), Hall, DiCiccio, and Romano (1989), Wang (1989), Lee and Young (1994), and Shao and Tu (1995). The smoothed bootstrap is an invaluable tool when testing for multi-modality in probability density estimators (Silverman, 1981; Silverman, 1983).

4.2.1 Failed Jackknife Calibration

The jackknife, introduced by Quenouille (1949), constructs n subsamples by sequentially deleting each datum from the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$. These subsamples are then used to construct an estimator of some population quantity. In this case, the population quantity is the level tuner for the nominal one-sided α-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}]$, where $\hat{\xi}_{\mathrm{Q2},\alpha}$ is given by (4.10), i.e. the level tuner β which satisfies

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q2},\beta}) = \alpha. \tag{4.34}$$

Set $\beta \in (0,1)$ and let $p_n(\beta) = P(X \in \widehat{\mathcal{I}}_{\mathrm{Q2},\beta})$, where the sample size n is explicitly referenced, and define the subsamples $\mathcal{X}_i = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)^\top$ for $i = 1, 2, \ldots, n$, of size $n-1$, which are obtained from the sample $\mathcal{X}$ by sequentially deleting the datum X_i. Denote by $\widehat{\mathcal{I}}_{\mathrm{Q2},i,\beta}$ the version of $\widehat{\mathcal{I}}_{\mathrm{Q2},\beta}$ computed using the subsample $\mathcal{X}_i$ of size $n-1$ instead of the sample $\mathcal{X}$ of size n.

The preliminary jackknife estimator of $p_n(\beta)$ is then given by

$$\hat{p}(\beta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{\widehat{\mathcal{I}}_{\mathrm{Q2},i,\beta}}(X_i). \tag{4.35}$$

Additionally, since the upper end-point of the nominal one-sided β-level prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q2},i,\beta}$ is calculated using the subsample $\mathcal{X}_i$ of size $n-1$, which is independent of X_i, it follows that

$$E\{\hat{p}(\beta)\} = p_{n-1}(\beta).$$

To remove the jumps present in the estimator given by (4.35), a linear interpolated form is introduced. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$, with $X_{i:n}$ denoting the i-th order statistic of the sample $\mathcal{X}$, and let

$$J_i(x) = \begin{cases} 1 & \text{if } x \geq X_{i+1:n}, \\ 0 & \text{if } x \leq X_{i:n}, \\ \frac{x - X_{i:n}}{X_{i+1:n} - X_{i:n}} & \text{if } X_{i:n} < x < X_{i+1:n}. \end{cases} \tag{4.36}$$

Since $X_{n:n} > \hat{\xi}_{\mathrm{Q2},n,\beta}$ for any $\beta \in (0,1)$, the former notation may be consistently extended by taking $J_n(\hat{\xi}_{\mathrm{Q2},n,\beta}) = 0$. The linearly interpolated form of (4.35) is then denoted by

$$\hat{p}_{\mathrm{J}}(\beta) = n^{-1}\sum_{i=1}^{n} J_i(\hat{\xi}_{\mathrm{Q2},i,\beta}).$$

The estimator of the level tuner β at (4.34), obtained using the linearly interpolated jackknife, is then given by the solution $\hat{\beta}$ of the equation

$$\hat{p}_{\mathrm{J}}(\beta) = \alpha.$$

The nominal one-sided α-level prediction interval

$$\widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\hat{\beta}}] \tag{4.37}$$

then represents the level-calibrated version of $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, derived using the linearly interpolated jackknife. However, this level-calibrated prediction interval fails to reduce coverage error. For the uncalibrated prediction interval it was shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}) = \alpha + O(n^{-2})$$

(Subsection 4.1.2), while for the level-calibrated prediction interval it can be shown that

$$P(X \leq \widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha}) = \alpha + O(n^{-1}). \tag{4.38}$$

Therefore, level calibration using the linearly interpolated jackknife increases coverage error from $O(n^{-2})$ to $O(n^{-1})$. Sufficient regularity conditions for (4.38) to hold are that F' exists, is Hölder continuous, and nonvanishing in a neighbourhood of ξ_α, the α-th population quantile. More precisely, it can be shown that

$$\limsup_{n\to\infty} n\left|P(X \in \widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha}) - \alpha\right| > 0. \tag{4.39}$$

Hence, the remainder given by (4.38) is the least possible. Results (4.38) and (4.39) are obtained from Theorem 4.3 of Section 4.3.

4.2.2 Smoothed Bootstrap Calibration

Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables from a population with F as probability distribution function. In Chapter 3 it has been seen that various estimators of population quantities were constructed using the bootstrap sample $\mathcal{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)^\top$ which, conditional on the sample $\mathcal{X}$, denotes n independent and identically distributed random variables from a population with $F_n(x) = n^{-1}\sum_{i=1}^n \mathbf{I}_{(-\infty,x]}(X_i)$ as probability distribution function. A perceived disadvantage that may vitiate performance of the estimator of the population quantity, and in turn the derived prediction interval, is that the bootstrap sample $\mathcal{X}^*$ may contain repeated data, or equivalently, data ties. By constructing a sample from a population which has a continuous probability distribution function, data ties are avoided. The smoothed bootstrap sample $\mathcal{X}^\dagger = (X_1^\dagger, X_2^\dagger, \ldots, X_n^\dagger)^\top$ denotes, conditional on the sample $\mathcal{X}$, n independent and identically distributed random variables from a population with kernel distribution estimator $\hat{F}$ as continuous probability distribution function. Therefore, with probability one, $\mathcal{X}^\dagger$ does not have data ties. For a critical comparison between the smoothed bootstrap and the bootstrap, or equivalently, the kernel distribution estimator $\hat{F}$ and the empirical distribution function F_n, see Silverman and Young (1987), Falk (1983), and Reiss (1981).

Additive calibration of the nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}]$ and $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q3},\alpha}]$, where $\hat{\xi}_{\mathrm{Q2},\alpha}$ and $\hat{\xi}_{\mathrm{Q3},\alpha}$ are given by (4.10) and (4.17) respectively, is

based on replacing additive tuners a_2 and a_3, which satisfy

$$P(X \in \mathcal{I}_{\mathrm{AC,Q2},\alpha}) = \alpha \quad \text{and} \quad P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}) = \alpha,$$

with $\mathcal{I}_{\mathrm{AC,Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha} + a_2]$ and $\mathcal{I}_{\mathrm{AC,Q3},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q3},\alpha} + a_3]$, by estimators $\hat{a}_2$ and $\hat{a}_3$, respectively. The procedure for generating estimates $\hat{a}_2$ and $\hat{a}_3$ using the smoothed bootstrap is as follows.

Suppose the population probability distribution function F has a probability density function denoted by f. Let $\hat{f}$ denote the kernel density estimator for f defined by

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{4.40}$$

where the kernel K is a probability density function and the bandwidth $h \to 0$ as $n \to \infty$. This density estimator was considered in Section 3.1 of Chapter 3 to construct an estimator for the α-th population quantile.

The kernel distribution estimator $\hat{F}$ for F then satisfies

$$\begin{aligned}
\hat{F}(x) &= \int_{-\infty}^{x} \hat{f}(y)\, dy \\
&= \int_{-\infty}^{x} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - X_i}{h}\right) dy \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{x} \frac{1}{h} K\left(\frac{y - X_i}{h}\right) dy \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\frac{x - X_i}{h}} K(z)\, dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{x-t}{h}} K(z)\, dz\, dF_n(t).
\end{aligned}$$

The smoothed bootstrap sample $\mathcal{X}^\dagger = (X_1^\dagger, X_2^\dagger, \ldots, X_n^\dagger)^\top$ denotes, conditional on the sample $\mathcal{X}$, n independent and identically distributed random variables drawn from a population with $\hat{F}$ as probability distribution function. The smoothed bootstrap predictand $X^\dagger$ denotes, conditional on the sample $\mathcal{X}$, a random variable which is independent of $\mathcal{X}^\dagger$ but drawn from the same population. The smoothed bootstrap estimators of

$$p_2(a_2) = P(X \in \mathcal{I}_{\mathrm{AC,Q2},\alpha}) \quad \text{and} \quad p_3(a_3) = P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha})$$

are given by

$$\hat{p}_2(a_2) = P(X^\dagger \in \mathcal{I}^\dagger_{\mathrm{AC,Q2},\alpha} \,|\, \mathcal{X}) \quad \text{and} \quad \hat{p}_3(a_3) = P(X^\dagger \in \mathcal{I}^\dagger_{\mathrm{AC,Q3},\alpha} \,|\, \mathcal{X}),$$

respectively, where $\hat{\xi}^{\dagger}_{Q2,\alpha}$ and $\hat{\xi}^{\dagger}_{Q3,\alpha}$ denote versions of $\hat{\xi}_{Q2,\alpha}$ and $\hat{\xi}_{Q3,\alpha}$, respectively, calculated using the smoothed bootstrap sample $\mathcal{X}^{\dagger}$ instead of the sample $\mathcal{X}$, and $\mathcal{I}^{\dagger}_{AC,Q2,\alpha} = (-\infty, \hat{\xi}^{\dagger}_{Q2,\alpha} + a_2]$ and $\mathcal{I}^{\dagger}_{AC,Q3,\alpha} = (-\infty, \hat{\xi}^{\dagger}_{Q3,\alpha} + a_3]$. Estimators of the additive tuners a_2 and a_3 obtained using the smoothed bootstrap are then given by the solutions $\hat{a}_2$ and $\hat{a}_3$ of the equations

$$\hat{p}_2(a_2) = \alpha \quad \text{and} \quad \hat{p}_3(a_3) = \alpha,$$

respectively.

The nominal one-sided α-level prediction intervals

$$\widehat{\mathcal{I}}_{AC,Q2,\alpha} = (-\infty, \hat{\xi}_{Q2,\alpha} + \hat{a}_2] \quad \text{and} \quad \widehat{\mathcal{I}}_{AC,Q3,\alpha} = (-\infty, \hat{\xi}_{Q3,\alpha} + \hat{a}_3] \tag{4.41}$$

then represent the additive-calibrated versions of $\widehat{\mathcal{I}}_{Q2,\alpha}$ and $\widehat{\mathcal{I}}_{Q3,\alpha}$, respectively, derived using the smoothed bootstrap. For the latter uncalibrated prediction intervals it was shown in Section 4.1 that

$$P(X \in \widehat{\mathcal{I}}_{Q2,\alpha}) = \alpha + O(n^{-2}),$$
$$P(X \in \widehat{\mathcal{I}}_{Q3,\alpha}) = \alpha + O(n^{-3}),$$

while for the former calibrated prediction intervals it can be shown that

$$\begin{aligned} P(X \in \widehat{\mathcal{I}}_{AC,Q2,\alpha}) &= \alpha + O(n^{-2-2/3}) \quad \text{if } h \sim cn^{-1/3}, \\ P(X \in \widehat{\mathcal{I}}_{AC,Q3,\alpha}) &= \alpha + O(n^{-3-2/5}) \quad \text{if } h \sim cn^{-1/5}, \end{aligned} \tag{4.42}$$

where $c > 0$ is any positive constant. Sufficient regularity conditions are that F has five bounded derivatives in a neighbourhood of the α-th population quantile ξ_α, that $f(\xi_\alpha) > 0$, that the kernel K is a symmetric and compactly supported probability density function with three bounded derivatives, and $E(|X|^l) < \infty$ for some l sufficiently large.

Result (4.42) is given by Theorem 4.4 of Section 4.3 and essentially relies on Taylor expansion and the Borel-Cantelli lemma conjoined with three inequalities, viz. Hoeffding's, Markov's, and Rosenthal's inequalities (see Section 4.3 for references).

The practical selection of the bandwidth h for a kernel density estimator is a point of contention. Various bandwidth selection techniques are given in Silverman (1986, Section 3.4). When the smoothed bootstrap is used for additive calibration of the nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{Q2,\alpha}$ and $\widehat{\mathcal{I}}_{Q3,\alpha}$ the bandwidth h may be trivially given by a constant multiple of $n^{-1/3}$

and $n^{-1/5}$, respectively. Alternatively, given a plausible parametric model for the population distribution function, e.g. a Normal population with estimated mean and variance, choose the bandwidth h such that it minimises absolute coverage error, and use this h as the bandwidth in the nonparametric procedure. A precedent for selecting bandwidth in an analogous way is given by Young (1988, Section 3).

To elucidate this further, let $f(x;\hat{\theta})$ denote the probability density function of a parametric population with estimated parameter vector $\hat{\theta} = \hat{\theta}(\mathcal{X})$ and let the parametric sample $\mathcal{X}' = (X_1', X_2', \ldots, X_n')^\top$ denote, conditional on $\mathcal{X}$, n independent and identically distributed random variables drawn from a population which has $f(x;\hat{\theta})$ as probability density function. Denote by $\widehat{\mathcal{I}}'_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}'_{\mathrm{AC,Q3},\alpha}$ the versions of $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively, constructed using the parametric sample $\mathcal{X}'$ instead of the sample $\mathcal{X}$. The bandwidths $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$ used to construct the additive-calibrated prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively, via the smoothed bootstrap can then be given by

$$\hat{h}_{\mathrm{P,Q2},\alpha} = \arg\min_h \left| P(X' \in \widehat{\mathcal{I}}'_{\mathrm{AC,Q2},\alpha} \mid \mathcal{X}) - \alpha \right|,$$
$$\hat{h}_{\mathrm{P,Q3},\alpha} = \arg\min_h \left| P(X' \in \widehat{\mathcal{I}}'_{\mathrm{AC,Q3},\alpha} \mid \mathcal{X}) - \alpha \right|,$$

where X' denotes, conditional on $\mathcal{X}$, a random variable which is independent of $\mathcal{X}'$ but drawn from the same population which has $f(x;\hat{\theta})$ as probability density function.

An alternative bandwidth procedure which obviates specification of a parametric model is as follows. Calculate bandwidths $\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{N,Q3},\alpha}$ which denote versions of $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$, respectively, where the parametric sample $\mathcal{X}'$ is drawn from a standard normal population, i.e. $f(x;\hat{\theta}) = (2\pi)^{-1/2}\exp(-x^2/2)$. Then the scaled quantities $\hat{h}_{\mathrm{S,Q2},\alpha} = \hat{\sigma}\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{S,Q3},\alpha} = \hat{\sigma}\hat{h}_{\mathrm{N,Q3},\alpha}$ are appropriate when constructing prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively, where $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

4.3 Theoretical Properties

4.3.1 Statement and Proof of Theorem 4.1

Theorem 4.1 *Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables which are drawn from a population with F as probability distribution function and $f = F'$ as probability density function. Assume that the α-th population quantile*

$\xi_\alpha = F^{-1}(\alpha) = \inf\{\xi : F(\xi) \geq \alpha\}$ is uniquely defined, that F has three bounded derivatives in a neighbourhood of ξ_α, and that $f(\xi_\alpha) > 0$. Given any $\alpha \in (0,1)$, select $2 \leq m \leq n-1$ (if possible) such that

$$|(n+1)\alpha - m| \leq \tfrac{2}{\sqrt{3}}, \tag{4.43}$$

and put $a = \frac{1}{2}(2-C\pm D)$, $b = -(1\pm D)$, $c = \frac{1}{2}(2+C\pm D)$, where $C = (n+1)\alpha - m$, $D = \sqrt{4-3C^2}$, and the $+$ or the $-$ sign are taken throughout in the definition of the triplet (a,b,c). Set

$$\hat{\xi}_{Q3,\alpha} = aX_{m-1:n} + bX_{m:n} + cX_{m+1:n}$$

and $\widehat{\mathcal{I}}_{Q3,\alpha} = (-\infty, \hat{\xi}_{Q3,\alpha}]$. Then

$$P(X \in \widehat{\mathcal{I}}_{Q3,\alpha}) = \alpha + O(n^{-3}),$$

where the predictand X denotes a random variable which is independent of the sample $\mathcal{X}$ but drawn from the same population.

Proof. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$, with $X_{i:n}$ denoting the i-th order statistic of the sample $\mathcal{X}$. Denote by $\mathcal{N}$ the neighbourhood of the α-th population quantile ξ_α in which the population distribution function F has three bounded derivatives, and select $\epsilon > 0$ sufficiently small such that

$$\{\xi : |\xi - \xi_\alpha| \leq \epsilon\} \subset \mathcal{N}.$$

Using a technique credited to Smirnov (1952) which employs properties of the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^n \mathbf{I}_{(-\infty,x]}(X_i)$ (Serfling, 1980; Reiss, 1989), observe that

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) = P(X_{m:n} > \xi_\alpha + \epsilon) + P(X_{m:n} < \xi_\alpha - \epsilon). \tag{4.44}$$

Consider the first term on the right hand side of (4.44). Then,

$$\begin{aligned}
P(X_{m:n} > \xi_\alpha + \epsilon) &= P\{F_n^{-1}(m/n) > \xi_\alpha + \epsilon\} \\
&= P\{m/n > F_n(\xi_\alpha + \epsilon)\} \\
&= P\Big\{m > n - \sum_{i=1}^n \mathbf{I}_{(\xi_\alpha+\epsilon,\infty)}(X_i)\Big\} \\
&= P\Big\{\sum_{i=1}^n \mathbf{I}_{(\xi_\alpha+\epsilon,\infty)}(X_i) > n(1-m/n)\Big\} \\
&= P\Big[\sum_{i=1}^n \{V_i - E(V_i)\} > n\{F(\xi_\alpha+\epsilon) - m/n\}\Big],
\end{aligned} \tag{4.45}$$

where $V_i = \mathbf{I}_{(\xi_\alpha+\epsilon,\infty)}(X_i)$.

Since m satisfies (4.43) it follows that

$$\frac{m}{n} = \alpha + O(n^{-1}). \tag{4.46}$$

Therefore, for n sufficiently large, a Taylor expansion argument can be used to show that $F(\xi_\alpha + \epsilon) - m/n > 0$. This allows the application of Hoeffding's inequality (Hoeffding, 1963; Serfling, 1980; Pollard, 1984) to (4.45). Hence, for n sufficiently large,

$$P(X_{m:n} > \xi_\alpha + \epsilon) \leq e^{-2nt_1^2}, \tag{4.47}$$

where $t_1 = F(\xi_\alpha + \epsilon) - m/n$.

Consider the second term on the right hand side of (4.44). Analogous to (4.45),

$$\begin{aligned} P(X_{m:n} < \xi_\alpha - \epsilon) &\leq P(X_{m:n} \leq \xi_\alpha - \epsilon) \\ &= P\Big[\sum_{i=1}^{n}\{V_i' - E(V_i')\} \geq n\{m/n - F(\xi_\alpha - \epsilon)\}\Big], \end{aligned} \tag{4.48}$$

where $V_i' = \mathbf{I}_{(\infty,\xi_\alpha-\epsilon]}(X_i)$. Since m satisfies (4.46) it follows that $m/n - F(\xi_\alpha - \epsilon) > 0$ which allows application of Hoeffding's inequality to (4.48). Hence, for n sufficiently large,

$$P(X_{m:n} < \xi_\alpha - \epsilon) \leq e^{-2nt_2^2}, \tag{4.49}$$

where $t_2 = m/n - F(\xi_\alpha - \epsilon)$.

Substituting bounds (4.47) and (4.49) into (4.44),

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) \leq 2e^{-2nt^2} = O(e^{-\rho n}), \tag{4.50}$$

where $t = \min\{t_1, t_2\}$ and $\rho = 2t^2 > 0$. From L'Hôpital's rule, replace (4.50) by

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.51}$$

for any $\lambda > 0$. This result is a consequence of, for n sufficiently large,

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) \leq P(|X_{m:n} - \xi_\alpha| > \epsilon_n), \tag{4.52}$$

where

$$\epsilon_n = \frac{1}{f(\xi_\alpha)}\sqrt{\frac{\lambda \log(n)}{n}}$$

for any $\lambda > 0$, since $\log(n)/n \to 0$ as $n \to \infty$.

How (4.51) follows from (4.52) can be elucidated as follows. By Taylor expansion and (4.46),

$$\rho_n n \geq \lambda \log(n), \tag{4.53}$$

where $\rho_n = 2t_n^2$ and $t_n = \min\{F(\xi_\alpha + \epsilon_n) - m/n, m/n - F(\xi_\alpha - \epsilon_n)\}$, for n sufficiently large. Hence, from (4.53) and by employing a version of (4.50) in which ϵ and ρ are replaced by ϵ_n and ρ_n respectively, observe that

$$P(|X_{m:n} - \xi_\alpha| > \epsilon_n) = O(n^{-\lambda}) \tag{4.54}$$

for any $\lambda > 0$. Therefore, (4.51) follows from (4.52) and (4.54).

Similarly it can be shown that

$$P(|X_{m-1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.55}$$

and

$$P(|X_{m+1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.56}$$

for any $\lambda > 0$.

Using the Borel-Cantelli lemma (Chung, 1974, Section 4.2) in conjunction with (4.54) it can be deduced that

$$X_{m:n} - \xi_\alpha = O\{(\log n/n)^{1/2}\}$$

with probability one.

Employing versions of (4.54), (4.55), and (4.56), and since assumed requisite conditions on the triplet (a, b, c) are that $a + b + c = 1$, $a = O(1)$, $b = O(1)$, and $c = O(1)$, it can be shown that

$$\begin{aligned}
P(|\hat{\xi}_{\mathrm{Q3},\alpha} - \xi_\alpha| > \epsilon) &\leq P\{|a(X_{m-1:n} - \xi_\alpha)| + |b(X_{m:n} - \xi_\alpha)| + |c(X_{m+1:n} - \xi_\alpha)| > \epsilon\} \\
&\leq P\{|a(X_{m-1:n} - \xi_\alpha)| > \epsilon/3\} + P\{|b(X_{m:n} - \xi_\alpha)| > \epsilon/3\} \\
&\qquad + P\{|c(X_{m+1:n} - \xi_\alpha)| > \epsilon/3\} \\
&\leq P\{|(X_{m-1:n} - \xi_\alpha)| > \epsilon/(3C)\} + P\{|(X_{m:n} - \xi_\alpha)| > \epsilon/(3C)\} \\
&\qquad + P\{|(X_{m+1:n} - \xi_\alpha)| > \epsilon/(3C)\} \\
&= O(n^{-\lambda})
\end{aligned} \tag{4.57}$$

for any $\lambda > 0$, where C is a sufficiently large real constant.

Let $\mathcal{A}_n = \mathcal{A}_{1,n} \cap \mathcal{A}_{2,n} \cap \mathcal{A}_{3,n} \cap \mathcal{A}_{4,n}$, where $\mathcal{A}_{1,n} = \{X_{m-1:n} \in \mathcal{N}\}$, $\mathcal{A}_{2,n} = \{X_{m:n} \in \mathcal{N}\}$, $\mathcal{A}_{3,n} = \{X_{m+1:n} \in \mathcal{N}\}$, and $\mathcal{A}_{4,n} = \{\hat{\xi}_{Q3,\alpha} \in \mathcal{N}\}$. Using Boole's inequality in conjunction with (4.54), (4.55), (4.56), and (4.57) it follows that

$$P(\mathcal{A}_n^c) = P(\cup_{i=1}^4 \mathcal{A}_{i,n}^c) \leq \sum_{i=1}^{n} P(\mathcal{A}_{i,n}^c) = O(n^{-\lambda}) \tag{4.58}$$

for any $\lambda > 0$.

Let $\mathcal{U} = (U_1, U_2, \ldots, U_n)^\top$ denote n independent and identically distributed random variables from a population with $F_U(x) = x\mathbf{I}_{(0,1]}(x) + \mathbf{I}_{(1,\infty)}(x)$ as probability distribution function, and let $U_{1:n} < U_{2:n} < \cdots < U_{n:n}$, with $U_{i:n}$ denoting the i-th order statistic of $\mathcal{U}$. A consequence of (4.58) is that, without loss of generality, the population probability distribution function, F, can be assumed to be continuous with three bounded derivatives on the whole real line. Therefore it follows (Reiss, 1989, Theorem 1.2.5) that

$$(U_{1:n}, U_{2:n}, \ldots, U_{n:n}) \stackrel{d}{=} (F(X_{1:n}), \ldots, F(X_{n:n}))$$

and

$$(F^{-1}(U_{1:n}), F^{-1}(U_{2:n}), \ldots, F^{-1}(U_{n:n})) \stackrel{d}{=} (X_{1:n}, X_{2:n}, \ldots, X_{n:n}),$$

where $\stackrel{d}{=}$ denotes equality of random variables in distribution.

In the remainder of this section, replace $\stackrel{d}{=}$ by $=$ as a tacit assumption where appropriate. Then, since $X_{m+1:n} = F^{-1}(U_{m+1:n})$, it follows from the Inverse Function Theorem and Taylor expansion

that

$$\begin{aligned}
X_{m+1:n} &= F^{-1}(U_{m:n} + U_{m+1:n} - U_{m:n}) \\
&= F^{-1}(U_{m:n}) + (U_{m+1:n} - U_{m:n}) f(F^{-1}(U_{m:n}))^{-1} \\
&\quad - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(F^{-1}(U_{m:n})) f(F^{-1}(U_{m:n}))^{-3} \\
&\quad + \frac{1}{6}(U_{m+1:n} - U_{m:n})^3 \left[3\{f^{(1)}(F^{-1}(\bar{U}))\}^2 f(F^{-1}(\bar{U}))^{-5} - f^{(2)}(F^{-1}(\bar{U})) f(F^{-1}(\bar{U}))^{-4}\right] \\
&= X_{m:n} + (U_{m+1:n} - U_{m:n}) f(X_{m:n})^{-1} \\
&\quad - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(X_{m:n}) f(X_{m:n})^{-3} \\
&\quad + \frac{1}{6}(U_{m+1:n} - U_{m:n})^3 \left[3\{f^{(1)}(\bar{X})\}^2 f(\bar{X})^{-5} - f^{(2)}(\bar{X}) f(\bar{X})^{-4}\right] \\
&= X_{m:n} + (U_{m+1:n} - U_{m:n}) f(X_{m:n})^{-1} \\
&\quad - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(X_{m:n}) f(X_{m:n})^{-3} + O_p(n^{-3}),
\end{aligned} \tag{4.59}$$

where $f^{(j)}(x)$ for $j = 1, 2$ denotes the j-th derivative of $f(x)$, $\bar{U}$ is on the line segment between $U_{m:n}$ and $U_{m+1:n}$, and $\bar{X}$ is on the line segment between $X_{m:n}$ and $X_{m+1:n}$.

Note that a requisite condition for application of the Inverse Function Theorem is that $f(\xi_\alpha) \neq 0$, which in this case is satisfied by assumption. The $O_p(n^{-3})$ term at (4.59) is derived from the fact that

$$U_{m+1:n} - U_{m:n} = O_p(n^{-1}), \tag{4.60}$$

with the other relevant factor being bounded on $\mathcal{A}_n$. Verification of (4.60) follows from $E(U_{r:n}) = r/(n+1)$; hence $E(U_{m+1:n} - U_{m:n}) = 1/(n+1) < 1/n = O(n^{-1})$. Therefore, for every $\epsilon > 0$ let $B_\epsilon = 1/\epsilon$. Observe from Markov's inequality that

$$P(n|U_{m+1:n} - U_{m:n}| > B_\epsilon) \leq nE(|U_{m+1:n} - U_{m:n}|)/B_\epsilon < \epsilon.$$

The result follows.

Similarly, since $X_{m-1:n} = F(U_{m-1:n})$, it follows that

$$\begin{aligned} X_{m-1:n} &= F^{-1}(U_{m:n} + U_{m-1:n} - U_{m:n}) \\ &= X_{m:n} + (U_{m+1:n} - U_{m:n}) f(X_{m:n})^{-1} \\ &\quad - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(X_{m:n}) f(X_{m:n})^{-3} \\ &\quad + \frac{1}{6}(U_{m+1:n} - U_{m:n})^3 \left[3\{f^{(1)}(\check{X})\}^2 f(\check{X})^{-5} - f^{(2)}(\check{X}) f(\check{X})^{-4}\right] \\ &= X_{m:n} + (U_{m+1:n} - U_{m:n}) f(X_{m:n})^{-1} \\ &\quad - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(X_{m:n}) f(X_{m:n})^{-3} + O_p(n^{-3}), \end{aligned} \tag{4.61}$$

where $\check{X}$ is on the line segment between $X_{m-1:n}$ and $X_{m:n}$.

Since $a + b + c = 1$, and by Taylor expansion,

$$\begin{aligned} F(\hat{\xi}_{\mathrm{Q3},\alpha}) &= F\{X_{m:n} + a(X_{m-1:n} - X_{m:n}) + c(X_{m+1:n} - X_{m:n})\} \\ &= F(X_{m:n}) + \{a(X_{m-1:n} - X_{m:n}) + c(X_{m+1:n} - X_{m:n})\} f(X_{m:n}) \\ &\quad + \frac{1}{2}\{a(X_{m-1:n} - X_{m:n}) + c(X_{m+1:n} - X_{m:n})\}^2 f^{(1)}(X_{m:n}) \\ &\quad + \frac{1}{6}\{a(X_{m-1:n} - X_{m:n}) + c(X_{m+1:n} - X_{m:n})\}^3 f^{(2)}(\tilde{X}), \end{aligned} \tag{4.62}$$

where $\tilde{X}$ is on the line segment between $X_{m:n}$ and $\hat{\xi}_{\mathrm{Q3},\alpha}$. Substituting (4.59) and (4.61) into (4.62), observe that

$$\begin{aligned} F(\hat{\xi}_{\mathrm{Q3},\alpha}) &= U_{m:n} + a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n}) \\ &\quad + \frac{1}{2}\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 \\ &\quad - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f^{(1)}(X_{m:n}) f(X_{m:n})^{-2} + R_n, \end{aligned} \tag{4.63}$$

where

$$
\begin{aligned}
R_n = \frac{a}{6}(U_{m-1:n} - U_{m:n})^3 G^{(3)}(\tilde{X}_1) f(X_{m:n}) \\
+ \frac{c}{6}(U_{m+1:n} - U_{m:n})^3 G^{(3)}(\tilde{X}_2) f(X_{m:n}) \\
- \frac{a^2}{2}(U_{m-1:n} - U_{m:n})^3 G^{(1)}(X_{m:n}) G^{(2)}(\tilde{X}_3) f^{(1)}(X_{m:n}) \\
- \frac{ac}{2}(U_{m-1:n} - U_{m:n})^2 (U_{m+1:n} - U_{m:n}) G^{(2)}(\tilde{X}_3) G^{(1)}(X_{m:n}) f^{(1)}(X_{m:n}) \\
- \frac{c^2}{2}(U_{m+1:n} - U_{m:n})^3 G^{(1)}(X_{m:n}) G^{(2)}(\tilde{X}_4) f^{(1)}(X_{m:n}) \\
- \frac{ac}{2}(U_{m-1:n} - U_{m:n})(U_{m+1:n} - U_{m:n})^2 G^{(1)}(X_{m:n}) G^{(2)}(\tilde{X}_4) f^{(1)}(X_{m:n}) \\
+ \frac{ac}{4}(U_{m-1:n} - U_{m:n})^2 (U_{m+1:n} - U_{m:n})^2 G^{(2)}(\tilde{X}_3) G^{(2)}(\tilde{X}_4) f^{(1)}(X_{m:n}) \\
+ \frac{a^2}{8}(U_{m-1:n} - U_{m:n})^4 G^{(2)}(\tilde{X}_3)^2 f^{(1)}(X_{m:n}) \\
+ \frac{c^2}{8}(U_{m+1:n} - U_{m:n})^4 G^{(2)}(\tilde{X}_4)^2 f^{(1)}(X_{m:n}) \\
+ \frac{1}{6}\{a(U_{m-1:n} - U_{m:n}) G^{(1)}(\tilde{X}_5) \\
+ c(U_{m+1:n} - U_{m:n}) G^{(1)}(\tilde{X}_6)\}^3 f^{(2)}(\tilde{X}_7),
\end{aligned}
$$

$G^{(1)}(x) = f(x)^{-1}$, $G^{(2)}(x) = -f^{(1)}(x)f(x)^{-3}$, $G^{(3)}(x) = 3\{f^{(1)}(x)\}^2 f(x)^{-5} - f^{(2)}(x)f(x)^{-4}$, and $\tilde{X}_1$, $\tilde{X}_2$, $\tilde{X}_3$, $\tilde{X}_4$, $\tilde{X}_5$, $\tilde{X}_6$, $\tilde{X}_7$ are on the line segment between either $X_{m-1:n}$ and $X_{m:n}$, $X_{m+1:n}$ and $X_{m:n}$ or $\hat{\xi}_{\mathrm{Q3},\alpha}$ and $X_{m:n}$.

Since $U_{m:n} - U_{m-1:n} = O_p(n^{-1})$, $U_{m+1:n} - U_{m:n} = O_p(n^{-1})$, and $O_p(n^{-3}) + O_p(n^{-4}) = O_p(n^{-3})$, it can be shown that

$$
R_n = O_p(n^{-3}).
$$

Since a version of (4.63) holds on $\mathcal{A}_n$,

$$
\begin{aligned}
E\{F(\hat{\xi}_{\mathrm{Q3},\alpha})\} = E\{F(\hat{\xi}_{\mathrm{Q3},\alpha})(\mathbf{I}_{\mathcal{A}_n} + \mathbf{I}_{\mathcal{A}_n^c})\} = E\Big(U_{m:n}\mathbf{I}_{\mathcal{A}_n} + a(U_{m-1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n} \\
+ c(U_{m+1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n} + \frac{1}{2}\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 \\
- a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f'(X_{m:n}) f(X_{m:n})^{-2} \mathbf{I}_{\mathcal{A}_n} \\
+ R_n \mathbf{I}_{\mathcal{A}_n} + F(\hat{\xi}_{\mathrm{Q3},\alpha})\mathbf{I}_{\mathcal{A}_n^c}\Big). \qquad (4.64)
\end{aligned}
$$

Using the tower property of conditional expectation (see (2.12)) with $\mathcal{G}_1 = \{\emptyset, \Omega\}$, $\mathcal{G}_2 = \sigma(\mathcal{X})$,

observe that

$$\begin{aligned} E\{F(\hat{\xi}_{\mathrm{Q3},\alpha})\} &= E\big[E\{\mathbf{I}_{(-\infty,\hat{\xi}_{\mathrm{Q3},\alpha}]}(X) \mid \mathcal{X}\}\big] \\ &= E\{\mathbf{I}_{(-\infty,\hat{\xi}_{\mathrm{Q3},\alpha}]}(X)\} \\ &= P(X \in \widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}), \end{aligned} \tag{4.65}$$

where X denotes a random variable which is independent of the sample $\mathcal{X}$ but drawn from the same population. Therefore, identifying bounds for (4.64) will automatically deliver appropriate bounds for the coverage error of $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$.

Since $F \leq 1$, use (4.58) directly to obtain

$$E\{F(\hat{\xi}_{\mathrm{Q3},\alpha})\mathbf{I}_{\mathcal{A}_n^c}\} \leq P(\mathcal{A}_n^c) = O(n^{-\lambda_1}), \tag{4.66}$$

for any $\lambda_1 > 0$.

To identify and bound the components of the summands at (4.64), replace $\mathbf{I}_{\mathcal{A}_n}$ by $1 - \mathbf{I}_{\mathcal{A}_n^c}$. Then

$$E(U_{m:n}\mathbf{I}_{\mathcal{A}_n}) = E(U_{m:n}) - E(U_{m:n}\mathbf{I}_{\mathcal{A}_n^c}), \tag{4.67}$$

$$E\{a(U_{m-1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n}\} = E\{a(U_{m-1:n} - U_{m:n})\} - E\{a(U_{m-1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n^c}\}, \tag{4.68}$$

and

$$E\{c(U_{m+1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n}\} = E\{c(U_{m+1:n} - U_{m:n})\} - E\{c(U_{m+1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n^c}\}. \tag{4.69}$$

Proceed by bounding the terms containing $\mathbf{I}_{\mathcal{A}_n^c}$ in (4.67), (4.68), and (4.69).

Using the fact that $U_{r:n} \leq 1$, and employing (4.58) directly, it can be shown that

$$E(U_{m:n}\mathbf{I}_{\mathcal{A}_n^c}) = O(n^{-\lambda_2}) \tag{4.70}$$

for any $\lambda_2 > 0$.

Note that the coefficients a, b, c are bounded in n and that $U_{m:n} - U_{m-1:n}$ has a Beta distribution which does not depend on m (David, 1981). Using these properties, and Jensen's and Hölder's inequalities in conjunction with (4.58), it can be shown that

$$\begin{aligned} \left|E\{a(U_{m-1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n^c}\}\right| &\leq E\{|a(U_{m-1:n} - U_{m:n})|\,|\mathbf{I}_{\mathcal{A}_n^c}|\} \\ &\leq E\{|a(U_{m-1:n} - U_{m:n})|^2\}^{1/2} P(\mathcal{A}_n^c)^{1/2} \\ &= O(n^{-1})O(n^{-\lambda_3/2}) \\ &= O(n^{-1-\lambda_3/2}) \end{aligned} \tag{4.71}$$

for any $\lambda_3 > 0$. Similarly,

$$E\{c(U_{m+1:n} - U_{m:n})\mathbf{I}_{\mathcal{A}_n^c}\} = O(n^{-1-\lambda_4/2}) \tag{4.72}$$

and

$$E(R_n\mathbf{I}_{\mathcal{A}_n}) = O(n^{-3}) \tag{4.73}$$

for any $\lambda_4 > 0$.

Take $\lambda_1 \geq 3$, $\lambda_2 \geq 3$, $\lambda_3 \geq 4$, and $\lambda_4 \geq 4$ in (4.66), (4.70), (4.71), and (4.72), respectively; repeatedly use the fact that $O(a_n) + O(b_n) = O(\max\{a_n, b_n\})$; use the bounds given by (4.65) and (4.73) together with

$$E(U_{r:n}) = \frac{r}{n+1};$$

then observe that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}) = \frac{m+c-a}{n+1} + \frac{1}{2}E\Big(\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f'(X_{m:n})f(X_{m:n})^{-2}\mathbf{I}_{\mathcal{A}_n}\Big) + O(n^{-3}). \tag{4.74}$$

Let $g(U_{m:n}) = f'\{F^{-1}(U_{m:n})\}f\{F^{-1}(U_{m:n})\}^{-2}$. Note that F has three bounded derivatives in a neighbourhood, $\mathcal{N}$, of the population quantile ξ_α. Using Taylor expansion and the Inverse Function Theorem it can be deduced that

$$f'(X_{m:n})f(X_{m:n})^{-2} = g(\alpha) + g'\{\alpha + \theta(U_{m:n} - \alpha)\}(U_{m:n} - \alpha), \tag{4.75}$$

where $0 < \theta < 1$ and

$$g'(x) = f^{(2)}(F^{-1}(x))f(F^{-1}(x))^{-3} - 2f^{(1)}(F^{-1}(x))f(F^{-1}(x))^{-4}.$$

Replace $f^{(1)}(X_{m:n})f(X_{m:n})^{-2}$ by (4.75) in (4.74). Bounding the remainder term in a similar way to (4.71) it can be shown that

$$E\Big(\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f^{(1)}(\xi_\alpha)f(\xi_\alpha)^{-2}\mathbf{I}_{\mathcal{A}_n}\Big) = E\Big(\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f^{(1)}(\xi_\alpha)f(\xi_\alpha)^{-2}\Big) + O(n^{-2-\lambda_5/2})$$

for any $\lambda_5 > 0$. Since (Reiss, 1989, Section 1.7)

$$E(U^2_{m-1:n}) = \frac{m(m-1)}{(n+1)(n+2)}, \qquad E(U_{m-1:n}U_{m:n}) = \frac{(m-1)(m+1)}{(n+1)(n+2)},$$
$$E(U^2_{m:n}) = \frac{(m+1)m}{(n+1)(n+2)}, \qquad E(U^2_{m+1:n}) = \frac{(m+2)(m+1)}{(n+1)(n+2)},$$
$$E(U_{m:n}U_{m+1:n}) = \frac{m(m+2)}{(n+1)(n+2)}, \qquad E(U_{m-1:n}U_{m+1:n}) = \frac{(m-1)(m+2)}{(n+1)(n+2)},$$

it can be shown that when $\lambda_5 \geq 2$,

$$E\Big(\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] f^{(1)}(\xi_\alpha) f(\xi_\alpha)^{-2}\Big) = \frac{1}{n^2}\{(a-c)^2 + (a^2+c^2) - 2(a+c)\} f^{(1)}(\xi_\alpha) f(\xi_\alpha)^{-2} + O(n^{-3}). \quad (4.76)$$

Furthermore (Reiss, 1989, Section 1.7),

$$E(U^2_{m-1:n}U_{m:n}) = \frac{m(m-1)(m+2)}{(n+1)(n+2)(n+3)}, \qquad E(U_{m-1:n}U^2_{m:n}) = \frac{(m-1)(m+1)(m+2)}{(n+1)(n+2)(n+3)},$$
$$E(U^3_{m:n}) = \frac{(m+2)(m+1)m}{(n+1)(n+2)(n+3)}, \qquad E(U_{m:n}U^2_{m+1:n}) = \frac{m(m+3)(m+2)}{(n+1)(n+2)(n+3)},$$
$$E(U^2_{m:n}U_{m+1:n}) = \frac{m(m+1)(m+3)}{(n+1)(n+2)(n+3)}, \qquad E(U_{m-1:n}U_{m:n}U_{m+1:n}) = \frac{(m-1)(m+1)(m+3)}{(n+1)(n+2)(n+3)}.$$

Using these results, and employing (4.43), it can be shown that

$$E\Big(\Big[\{a(U_{m-1:n} - U_{m:n}) + c(U_{m+1:n} - U_{m:n})\}^2 - a(U_{m-1:n} - U_{m:n})^2 - c(U_{m+1:n} - U_{m:n})^2\Big] g'\{\alpha + \theta(U_{m:n} - \alpha)\}(U_{m:n} - \alpha)\mathbf{I}_{\mathcal{A}_n}\Big) = O(n^{-3}). \quad (4.77)$$

Substituting (4.76) and (4.77) into (4.74) it can be shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}) = \frac{m+c-a}{n+1} + \frac{1}{2n^2}\{(a-c)^2 + (a^2+c^2) - 2(a+c)\} f^{(1)}(\xi_\alpha) f(\xi_\alpha)^{-2} + O(n^{-3}). \quad (4.78)$$

The asymptotic expansion at (4.78) was derived under the assumption that $a + b + c = 1$ and that the triplet (a, b, c) depends on the sample size n and α only and is not dependent on the probability distribution function. Therefore requisite conditions for $P(X \in \widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}) = \alpha + O(n^{-3})$ are that

$$c - a = C$$

and

$$(a-c)^2 + a^2 + c^2 - 2(a+c) = 0,$$

where $C = (n+1)\alpha - m$.

Solving the two former equations for a it follows that $a = \frac{1}{2}(2 - C \pm D)$, where $D = \sqrt{4 - 3C^2}$. Then, since $a \in \mathbb{R}$, it is required that $4 - 3C^2 \geq 0$, or equivalently,

$$\left|m - (n+1)\alpha\right| \leq \frac{2}{\sqrt{3}}.$$

When $2/(n+1) \leq \alpha \leq (n-2)/(n+1)$, the former inequality is satisfied by at least one value of m in the range $2 \leq m \leq n-1$. Setting $c = C + a = \frac{1}{2}(2 + C \pm D)$ and $b = 1 - a - c = -(1 \pm D)$ finalises the proof. ∎

4.3.2 Statement and Proof of Theorem 4.2

Theorem 4.2 *Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables which are drawn from a population with F as probability distribution function and $f = F'$ as probability density function. Assume that the α-th population quantile $\xi_\alpha = F^{-1}(\alpha) = \inf\{\xi : F(\xi) \geq \alpha\}$ is uniquely defined, that F has four bounded derivatives in a neighbourhood of ξ_α, and that $f(\xi_\alpha) > 0$. Set*

$$\hat{\xi}_{4,\alpha} = aX_{m:n} + bX_{m+1:n} + cX_{m+2:n} + dX_{m+3:n}$$

and $\widehat{\mathcal{I}}_{4,\alpha} = (-\infty, \hat{\xi}_{4,\alpha}]$. Then for all $\alpha \in (0,1)$ it is not possible to select $a = O(1)$, $b = O(1)$, $c = O(1)$, $d = O(1)$, and $m = (n+1)\alpha + O(1)$, depending on the sample size n and α only, and satisfying $a + b + c + d = 1 + o(1)$, such that

$$P(X \in \widehat{\mathcal{I}}_{4,\alpha}) = \alpha + O(n^{-4}),$$

where the predictand X denotes a random variable which is independent of the sample $\mathcal{X}$ but drawn from the same population.

Proof. Let $\alpha \in (0,1)$ and denote by $\mathcal{N}$ the neighbourhood of the α-th population quantile ξ_α in which the population distribution function F has four bounded derivatives and select $\epsilon > 0$ sufficiently small such that

$$\{\xi : |\xi - \xi_\alpha| \leq \epsilon\} \subset \mathcal{N}.$$

Using Hoeffding's inequality and Taylor expansion (see Theorem 4.1) it can be shown that

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.79}$$

$$P(|X_{m+1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.80}$$

$$P(|X_{m+2:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.81}$$

$$P(|X_{m+3:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.82}$$

and

$$P(|\hat{\xi}_{4,\alpha} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.83}$$

for any $\lambda > 0$.

Let $\mathcal{A}_n = \mathcal{A}_{1,n} \cap \mathcal{A}_{2,n} \cap \mathcal{A}_{3,n} \cap \mathcal{A}_{4,n} \cap \mathcal{A}_{5,n}$, where $\mathcal{A}_{1,n} = \{X_{m:n} \in \mathcal{N}\}$, $\mathcal{A}_{2,n} = \{X_{m+1:n} \in \mathcal{N}\}$, $\mathcal{A}_{3,n} = \{X_{m+2:n} \in \mathcal{N}\}$, $\mathcal{A}_{4,n} = \{X_{m+3:n} \in \mathcal{N}\}$, and $\mathcal{A}_{5,n} = \{\hat{\xi}_{4,\alpha} \in \mathcal{N}\}$. Using Boole's inequality in conjunction with (4.79), (4.80), (4.81), (4.82), and (4.83) it follows that

$$P(\mathcal{A}_n^c) = O(n^{-\lambda}) \tag{4.84}$$

for any $\lambda > 0$.

A consequence of (4.84) is that, without loss of generality, the population probability distribution function, F, can be assumed to be continuous with four bounded derivatives on the whole real line. Therefore, by the Inverse Function Theorem and Taylor expansion, it can be deduced that

$$\begin{aligned} X_{m+k:n} - X_{m:n} &= F^{-1}(U_{m:n} + U_{m+k:n} - U_{m:n}) - F^{-1}(U_{m:n}) \\ &= \sum_{r=1}^{3} \frac{1}{r!}\delta_k^r G^{(r)}(X_{m:n}) + \frac{1}{4!}\delta_k^4 G^{(4)}(\tilde{X}_k), \qquad k = 1,2,3, \end{aligned} \tag{4.85}$$

where $f^{(j)}(x)$ for $j = 1,2,3$ denotes the j-th derivative of $f(x)$, $G = F^{-1}$, $G^{(1)}(x) = f(x)^{-1}$, $G^{(2)}(x) = -f^{(1)}(x)f(x)^{-3}$, $G^{(3)}(x) = 3\{f^{(1)}(x)\}^2 f(x)^{-5} - f^{(2)}(x)f(x)^{-4}$, $G^{(4)}(x) = -f^{(3)}(x)f(x)^{-5} + 10f^{(1)}(x)f^{(2)}(x)f(x)^{-6} - 15\{f^{(1)}(x)\}^3 f(x)^{-7}$, $\delta_k = U_{m+k:n} - U_{m:n}$, and $\tilde{X}_k$ lies on the line segment between $X_{m+k:n}$ and $X_{m:n}$.

By attempting to select the quintuplet (m, a, b, c, d), where $m = (n+1)\alpha + O(1)$, $a = O(1)$, $b = O(1)$, $c = O(1)$, $d = O(1)$, and $a + b + c + d = 1 + o(1)$, and which depends only on the sample size n and α, such that

$$\begin{aligned} P(X \in \hat{\mathcal{I}}_{4,\alpha}) &= E\{F(\hat{\xi}_{4,\alpha})\} \\ &= \alpha + O(n^{-4}), \end{aligned} \tag{4.86}$$

it follows that

$$a+b+c+d=1+O(n^{-4}); \tag{4.87}$$

a larger order would contradict (4.86). Additionally, using Taylor expansion, the term of order n^{-4} at (4.87) can be mapped into the term of order n^{-4} at (4.86). Therefore, without loss of generality, it can be assumed that $a+b+c+d=1$.

Write $\hat{\xi}_{4,\alpha}=X_{m:n}+\Delta$, where

$$\Delta=b(X_{m+1:n}-X_{m:n})+c(X_{m+2:n}-X_{m:n})+d(X_{m+3:n}-X_{m:n}).$$

Then, by Taylor expansion, observe that

$$F(\hat{\xi}_{4,\alpha})=F(X_{m:n})+\Delta f(X_{m:n})+\frac{1}{2}\Delta^2 f^{(1)}(X_{m:n})+\frac{1}{6}\Delta^3 f^{(2)}(X_{m:n})+\frac{1}{24}\Delta^4 f^{(3)}(\bar{X}), \tag{4.88}$$

where $\bar{X}$ lies on the line segment between $\hat{\xi}_{4,\alpha}$ and $X_{m:n}$. Substituting (4.85) into (4.88) it can be deduced that

$$\begin{aligned}
F(\check{\xi}_\alpha)=F(X_{m:n})&+b\delta_1+c\delta_2+d\delta_3\\
&+\frac{1}{2}\{(b\delta_1+c\delta_2+d\delta_3)^2-(b\delta_1^2+c\delta_2^2+d\delta_3^2)\}f^{(1)}(X_{m:n})f(X_{m:n})^{-2}\\
&+\frac{1}{6}\Big[(b\delta_1+c\delta_2+d\delta_3)^3 f^{(2)}(X_{m:n})f(X_{m:n})^{-3}\\
&-3(b\delta_1+c\delta_2+d\delta_3)(b\delta_1^2+c\delta_2^2+d\delta_3^2)f^{(1)}(X_{m:n})^2 f(X_{m:n})^{-4}\\
&+(b\delta_1^3+c\delta_2^3+d\delta_3^3)\{3f^{(1)}(X_{m:n})^2 f(X_{m:n})^{-4}-f^{(2)}(X_{m:n})f(X_{m:n})^{-3}\}\Big]+R_n,
\end{aligned} \tag{4.89}$$

where R_n includes the remainder terms from the expansion.

Since a version of (4.89) holds on $\mathcal{A}_n$,

$$\begin{aligned}
P(X\in\widehat{\mathcal{I}}_{4,\alpha})&=E\{F(\hat{\xi}_{4,\alpha})(\mathbf{I}_{\mathcal{A}_n}+\mathbf{I}_{\mathcal{A}_n^c})\}\\
&=E\Big(U_{m:n}\mathbf{I}_{\mathcal{A}_n}+(b\delta_1+c\delta_2+d\delta_3)\mathbf{I}_{\mathcal{A}_n}\\
&\quad+\frac{1}{2}\{(b\delta_1+c\delta_2+d\delta_3)^2-(b\delta_1^2+c\delta_2^2+d\delta_3^2)\}f^{(1)}(X_{m:n})f(X_{m:n})^{-2}\mathbf{I}_{\mathcal{A}_n}\\
&\quad+\frac{1}{6}\Big[(b\delta_1+c\delta_2+d\delta_3)^3 f^{(2)}(X_{m:n})f(X_{m:n})^{-3}\\
&\quad-3(b\delta_1+c\delta_2+d\delta_3)(b\delta_1^2+c\delta_2^2+d\delta_3^2)f^{(1)}(X_{m:n})^2 f(X_{m:n})^{-4}\\
&\quad+(b\delta_1^3+c\delta_2^3+d\delta_3^3)\{3f^{(1)}(X_{m:n})^2 f(X_{m:n})^{-4}\\
&\qquad-f^{(2)}(X_{m:n})f(X_{m:n})^{-3}\}\Big]\mathbf{I}_{\mathcal{A}_n}+R_n\mathbf{I}_{\mathcal{A}_n}+F(\hat{\xi}_{4,\alpha})\mathbf{I}_{\mathcal{A}_n^c}\Big).
\end{aligned} \tag{4.90}$$

Since $F \leq 1$ it is possible to use (4.84) directly, for an appropriate $\lambda > 0$, to deduce that

$$E\{F(\hat{\xi}_{4,\alpha})\mathbf{I}_{\mathcal{A}_n^c}\} = O(n^{-4}). \tag{4.91}$$

Using moment properties of uniformly distributed order statistics it can be shown that

$$E(R_n\mathbf{I}_{\mathcal{A}_n^c}) = O(n^{-4}). \tag{4.92}$$

Additionally, from David (1981, Theorem 2.7) or Reiss (1989, Theorem 1.8.1), conditional on $X_{m:n}$ the random vector $(\delta_1/\{1-F(X_{m:n})\}, \delta_2/\{1-F(X_{m:n})\}, \delta_3/\{1-F(X_{m:n})\})^\top$ has the same probability distribution function as $(U_{1:n-m}, U_{2:n-m}, U_{3:n-m})^\top$. Denote $U_{k:n-m}$ by V_k for $k = 1, 2, 3$ which may be taken to be independent of $X_{m:n}$ (Chow and Teicher, 1997, Section 6.3). Note that $E(\delta_k) = k/(n+1)$ and that δ_k for $k = 1, 2, 3$ has a Beta distribution with parameters k and $n-k+1$ (David, 1981, Example 2.3).

Using (4.84), (4.91), (4.92), and Hölder's and Young's inequalities in conjunction with (4.90) it can be deduced that

$$\begin{aligned} P(X \in \widehat{\mathcal{I}}_{4,\alpha}) = {} & \frac{m+b+2c+3d}{n+1} + \frac{1}{2}\gamma_1(n)E\{(bV_1+cV_2+dV_3)^2 - (bV_1^2+cV_2^2+dV_3^2)\} \\ & + \frac{1}{6}\big[3\gamma_2(n)E\{(bV_1^3+cV_2^3+dV_3^3) - (bV_1+cV_2+dV_3)(bV_1^2+cV_2^2+dV_3^2)\} \\ & + \gamma_3(n)E\{(bV_1+cV_2+dV_3)^3 - (bV_1^3+cV_2^3+dV_3^3)\} + O(n^{-4}), \end{aligned} \tag{4.93}$$

where the terms with leading coefficients '1/2' and '1/6' on the right-hand side of (4.93) are of orders n^{-2} and n^{-3}, respectively, and

$$\begin{aligned} \gamma_1(n) &= E\big[f^{(1)}(X_{m:n})f(X_{m:n})^{-2}\{1-F(X_{m:n})\}^2\mathbf{I}_{\mathcal{A}_{1,n}}\big], \\ \gamma_2(n) &= E\big[f^{(1)}(X_{m:n})^2f(X_{m:n})^{-4}\{1-F(X_{m:n})\}^3\mathbf{I}_{\mathcal{A}_{1,n}}\big], \end{aligned}$$

and

$$\gamma_3(n) = E\big[f^{(2)}(X_{m:n})f(X_{m:n})^{-3}\{1-F(X_{m:n})\}^3\mathbf{I}_{\mathcal{A}_{1,n}}\big].$$

Since the quintuplet (m, a, b, c, d) depends on the sample size n and α only and not on the population distribution function, it follows that requisite conditions for (4.86) to be satisfied are

that

$$\frac{m+b+2c+3d}{n+1} = \alpha, \tag{4.94}$$

$$E\{(bV_1+cV_2+dV_3)^2-(bV_1^2+cV_2^2+dV_3^2)\}=0, \tag{4.95}$$

$$E\{(bV_1^3+cV_2^3+dV_3^3)-(bV_1+cV_2+dV_3)(bV_1^2+cV_2^2+dV_3^2)\}=0, \tag{4.96}$$

$$E\{(bV_1+cV_2+dV_3)^3-(bV_1^3+cV_2^3+dV_3^3)\}=0. \tag{4.97}$$

Since (Reiss, 1989, Section 1.7)

$$E(V_1^2)=\frac{2}{(n-m+1)(n-m+2)}, \quad E(V_1V_2)=\frac{3}{(n-m+1)(n-m+2)},$$
$$E(V_1V_3)=\frac{4}{(n-m+1)(n-m+2)}, \quad E(V_2^2)=\frac{6}{(n-m+1)(n-m+2)},$$
$$E(V_2V_3)=\frac{8}{(n-m+1)(n-m+2)}, \quad E(V_3^2)=\frac{12}{(n-m+1)(n-m+2)},$$

and

$$E(V_1^3)=6K, \quad E(V_1^2V_2)=8K, \quad E(V_1^2V_3)=10K, \quad E(V_1V_2^2)=12K, \quad E(V_2^2V_3)=30K,$$
$$E(V_2V_3^2)=40K, \quad E(V_1V_3^2)=20K, \quad E(V_1V_2V_3)=15K, \quad E(V_2^3)=24K, \quad E(V_3^3)=120K,$$

where $K=1/\{(n-m+1)(n-m+2)(n-m+3)\}$, it is possible to rewrite (4.94), (4.95), (4.96), and (4.97) in the equivalent forms

$$b+2c+3d=K_{n,\alpha,m}, \tag{4.98}$$

$$2(b^2-b)+6(c^2+c)+12(d^2-d)+6bc+8bd+16cd=0, \tag{4.99}$$

$$6(b-b^2)+24(c-c^2)+120(d-d^2)-20bc-30bd-70cd=0, \tag{4.100}$$

$$6(b^3-b)+24(c^3-c)+120(d^3-d)+24b^2c+30b^2d+36c^2b$$
$$+90c^2d+120cd^2+60bd^2+90bcd=0, \tag{4.101}$$

where $K_{n,\alpha,m}=(n+1)\alpha-m$.

Suppose there does exist a quintuplet (m,a,b,c,d) which satisfies (4.98), (4.99), (4.100), and (4.101). Solving (4.98) and (4.99) for c it can be shown that

$$c=\frac{(2+2K_{n,\alpha,m}-6d)\pm\sqrt{132d^2+(8K_{n,\alpha,m}+24)d+(4+23K_{n,\alpha,m}-12K_{n,\alpha,m}^2)}}{4}.$$

Since the unknown quantity $c\in\mathbb{R}$ it is required that

$$132d^2+(8K_{n,\alpha,m}+24)d+(4+23K_{n,\alpha,m}-12K_{n,\alpha,m}^2)\geq 0,$$

or equivalently, that the unknown quantity $d \in (-\infty, d_{1,-}] \cup [d_{1,+}, \infty)$, where

$$d_{1,\pm} = \frac{-8K_{n,\alpha,m} - 24 \pm \sqrt{6400K_{n,\alpha,m}^2 - 11760K_{n,\alpha,m} - 1536}}{264}.$$

Furthermore, since the unknown quantity $d \in \mathbb{R}$, it is required that

$$6400K_{n,\alpha,m}^2 - 11760K_{n,\alpha,m} - 1536 \geq 0,$$

or equivalently, that the unknown quantity m satisfies $K_{n,\alpha,m} \in (-\infty, K_{1,-}] \cup [K_{1,+}, \infty)$, where

$$K_{1,\pm} = \frac{147}{160} \pm \frac{29}{160}\sqrt{33}. \tag{4.102}$$

Solving (4.98) and (4.100) for c it can be shown that

$$c = \frac{\bar{e} \pm \sqrt{141044d^2 + (-13392 - 9584K_{n,\alpha,m})d + (144 - 864K_{n,\alpha,m} + 976K_{n,\alpha,m}^2)}}{80},$$

where $\bar{e} = (-12 - 4K_{n,\alpha,m} + 122d)$. Since the unknown quantity $c \in \mathbb{R}$ it follows that

$$141044d^2 + (-13392 - 9584K_{n,\alpha,m})d + (144 - 864K_{n,\alpha,m} + 976K_{n,\alpha,m}^2) \geq 0,$$

or equivalently, that the unknown quantity $d \in (-\infty, d_{2,-}] \cup [d_{2,+}, \infty)$, where

$$d_{2,\pm} = \frac{13392 + 9584K_{n,\alpha,m} \pm \sqrt{-458782720K_{n,\alpha,m}^2 + 744145920K_{n,\alpha,m} + 98104320}}{282088}.$$

Furthermore, since the unknown quantity $d \in \mathbb{R}$, it follows that

$$-458782720K_{n,\alpha,m}^2 + 744145920K_{n,\alpha,m} + 98104320 \geq 0,$$

or equivalently, that the unknown quantity m satisfies $K_{n,\alpha,m} \in [K_{2,-}, K_{2,+}]$, where

$$K_{2,\pm} = \frac{145341}{179212} \pm \frac{3}{179212}\sqrt{3110196505}. \tag{4.103}$$

Combining (4.102) and (4.103) it follows that the unknown quantity m satisfies

$$K_{n,\alpha,m} \in [K_{2,-}, K_{1,-}] \cup [K_{1,+}, K_{2,+}]. \tag{4.104}$$

However, when $(n+1)\alpha \in \mathbb{N}$ it is not possible to select an integer $1 \leq m \leq n-3$ such that (4.104) is satisfied. This delivers the required contradiction which concludes the proof. ∎

4.3.3 Statement and Proof of Theorem 4.3

Theorem 4.3 *Let the sample* $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ *denote* n *independent and identically distributed random variables which are drawn from a population with* F *as probability distribution function and* $f = F'$ *as probability density function. Assume that the* α*-th population quantile* $\xi_\alpha = F^{-1}(\alpha)$ *is uniquely defined and that* f *is Hölder continuous and nonvanishing in a neighbourhood of* ξ_α*. Let the nominal one-sided* α*-level prediction interval* $\widehat{\mathcal{I}}_{LC,Q2,\alpha}$ *be defined at* (4.37)*. Then*

$$\limsup_{n\to\infty} n\left|P(X \in \widehat{\mathcal{I}}_{LC,Q2,\alpha}) - \alpha\right| > 0,$$

where the predictand X *denotes a random variable which is independent of* $\mathcal{X}$ *but drawn from the same population.*

Proof. The nominal one-sided α-level prediction interval

$$\widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\hat{\beta}}]$$

was introduced in Subsection 4.2.1 and it represents the level-calibrated version of the nominal one-sided α-level prediction interval

$$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}]$$

defined at (4.11). Here the estimator of the level tuner is then given by the solution $\hat{\beta}$ of the equation

$$\hat{p}_{\mathrm{J}}(\beta) = \alpha,$$

where

$$\hat{p}_{\mathrm{J}}(\beta) = n^{-1}\sum_{i=1}^{n} J_i(\hat{\xi}_{\mathrm{Q2},i,\beta}),$$

$J_i(x)$ is defined at (4.36), and $J_n(\hat{\xi}_{\mathrm{Q2},n,\beta}) = 0$.

When constructing $\hat{\xi}_{\mathrm{Q2},i,\beta}$, which is a version of $\hat{\xi}_{\mathrm{Q2},\beta}$ defined at (4.10) and is constructed using the subsample $\mathcal{X}_i$ of size $n-1$ instead of the sample $\mathcal{X}$ of size n, it is necessary to replace $m = \lfloor (n+1)\alpha \rfloor$ (see (4.9)) by $m = \lfloor n\alpha \rfloor$. For ease of exposition the latter definition of m is used throughout this proof.

Note that $J_i(\hat{\xi}_{\mathrm{Q2},i,\beta}) = 1$ for $i = 1, 2, \ldots, m$ since $\hat{\xi}_{\mathrm{Q2},i,\beta} = (m+1-n\beta)X_{m+1:n} + (n\beta - m)X_{m+2:n}$ and $X_{i:n} \leq \hat{\xi}_{\mathrm{Q2},i,\beta}$. Additionally $J_i(\hat{\xi}_{\mathrm{Q2},i,\beta}) = 0$ for $i = m+2, m+3, \ldots, n-1$ since $\hat{\xi}_{\mathrm{Q2},i,\beta} =$

$(m+1-n\beta)X_{m:n} + (n\beta - m)X_{m+1:n}$ and $X_{i:n} > \hat{\xi}_{Q2,i,\beta}$. Since the m-th and $(m+1)$-th order statistics of the subsample $\mathcal{X}_{m+1}$ are $X_{m:n}$ and $X_{m+2:n}$, respectively, it follows that

$$\hat{\xi}_{Q2,m+1,\beta} = (m+1-n\beta)X_{m:n} + (n\beta - m)X_{m+2:n},$$

and hence

$$J_{m+1}(\hat{\xi}_{Q2,m+1,\beta}) = \max\left\{\frac{(n\beta - m)\big(X_{m+2:n} - X_{m:n}\big) - \big(X_{m+1:n} - X_{m:n}\big)}{X_{m+2:n} - X_{m+1:n}}, 0\right\}.$$

Therefore,

$$\hat{p}_{\mathrm{J}}(\beta) = \frac{m}{n} + \frac{1}{n}\max\left\{\frac{(n\beta - m)\big(X_{m+2:n} - X_{m:n}\big) - \big(X_{m+1:n} - X_{m:n}\big)}{X_{m+2:n} - X_{m+1:n}}, 0\right\}.$$

If it is assumed that

$$\frac{m}{n} + \frac{1}{n}\Big(\frac{X_{m+1:n} - X_{m:n}}{X_{m+2:n} - X_{m:n}}\Big) < \alpha < \frac{m+1}{n}$$

then

$$\max\left\{\frac{(n\alpha - m)\big(X_{m+2:n} - X_{m:n}\big) - \big(X_{m+1:n} - X_{m:n}\big)}{X_{m+2:n} - X_{m+1:n}}, 0\right\} > 0.$$

Hence the estimator of the level tuner derived using the linearly interpolated jackknife is given by the solution $\hat{\beta}$ of the equation $\hat{p}_{\mathrm{J}}(\beta) = \alpha$ and satisfies

$$\hat{\beta} = (m + \delta)/n,$$

where $\delta = (m+1-n\alpha)\Delta + (n\alpha - m)$ and $\Delta = \big(X_{m+1:n} - X_{m:n}\big)/\big(X_{m+2:n} - X_{m:n}\big)$.

Consider only elements of the sequence $\mathcal{S} = \mathcal{S}(\alpha, \eta)$ of values of n such that, for the definition $m = \lfloor n\alpha \rfloor$, $m/n < \alpha < (m+1-\alpha-\eta)/n$ for some fixed $\eta > 0$. The sequence $\mathcal{S}$ depends on α and η, and when $\alpha \in (0,1)$ and $\eta \in (0, 1-\alpha)$ it contains an infinite number of elements. For all $n \in \mathcal{S}$ sufficiently large, both $m/n < \alpha < (m+1)/n$ and $m/(n+1) < \alpha < (m+1)/(n+1)$. Hence it may be assumed without contradiction that $m = \lfloor (n+1)\alpha \rfloor$. Therefore,

$$\hat{\xi}_{Q2,\hat{\beta}} = X_{m:n} + \{(n+1)\hat{\beta} - m\}(X_{m+1:n} - X_{m:n}).$$

Denote by $\mathcal{N}$ the neighbourhood of the α-th population quantile ξ_α in which the population probability density function f is Hölder continuous and nonvanishing, and select $\epsilon > 0$ sufficiently small such that

$$\{\xi : |\xi - \xi_\alpha| \le \epsilon\} \subset \mathcal{N}.$$

Since $\hat{\beta} = \alpha + O(1)$ and by using Hoeffding's inequality and Taylor expansion (see Theorem 4.1) it can be shown that

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.105}$$

$$P(|X_{m+1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.106}$$

and

$$P(|\hat{\xi}_{\mathrm{Q2},\hat{\beta}} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.107}$$

for any $\lambda > 0$.

Let $\mathcal{A}_n = \mathcal{A}_{1,n} \cap \mathcal{A}_{2,n} \cap \mathcal{A}_{3,n}$, where $\mathcal{A}_{1,n} = \{X_{m:n} \in \mathcal{N}\}$, $\mathcal{A}_{2,n} = \{X_{m+1:n} \in \mathcal{N}\}$, and $\mathcal{A}_{3,n} = \{\hat{\xi}_{\mathrm{Q2},\hat{\beta}} \in \mathcal{N}\}$. Using Boole's inequality in conjunction with (4.105), (4.106), and (4.107) it follows that

$$P(\mathcal{A}_n^c) = O(n^{-\lambda}) \tag{4.108}$$

for any $\lambda > 0$.

A consequence of (4.108) is that, without loss of generality, the population probability density function f can be assumed to be Hölder continuous and nonvanishing on the whole real line. Therefore, by the Inverse Function Theorem and Taylor expansion, it can be deduced that

$$\begin{aligned}\Delta &= \frac{X_{m+1:n} - X_{m:n}}{X_{m+2:n} - X_{m:n}} \\ &= \frac{U_{m+1:n} - U_{m:n} - \frac{1}{2}(U_{m+1:n} - U_{m:n})^2 f^{(1)}(\tilde{X}) f(\tilde{X})^{-3} f(X_{m:n})^{-1}}{(U_{m+2:n} - U_{m:n})\{1 - \frac{1}{2}(U_{m+2:n} - U_{m:n}) f^{(1)}(\bar{X}) f(\bar{X})^{-3} f(X_{m:n})^{-1}},\end{aligned} \tag{4.109}$$

where $\tilde{X}$ and $\bar{X}$ are on the line segment between $X_{m+1:n}$ and $X_{m:n}$ and $X_{m+2:n}$ and $X_{m:n}$, respectively.

By Taylor expansion, observe that

$$\begin{aligned}F(\hat{\xi}_{\mathrm{Q2},\hat{\beta}}) = F(X_{m:n}) &+ \{(n+1)\hat{\beta} - m\}(X_{m+1:n} - X_{m:n}) f(X_{m:n}) \\ &+ \{(n+1)\hat{\beta} - m\}(X_{m+1:n} - X_{m:n})\{f(\check{X}) - f(X_{m:n})\},\end{aligned} \tag{4.110}$$

where $\check{X}$ is on the line segment between $\hat{\xi}_{\mathrm{Q2},\hat{\beta}}$ and $X_{m:n}$.

Since $\Delta < 1$ and the population probability density function f is Hölder continuous both Hölder's and Lyapunov's inequalities can be employed to derive:

$$\begin{aligned}
&\Big|E\big[\mathbf{I}_{\mathcal{A}_n}\{(n+1)\hat{\alpha}-m\}(X_{m+1:n}-X_{m:n})\{f(\check{X})-f(X_{m:n})\}\big]\Big| \\
&\qquad \le C_1 E\Big[\big|\mathbf{I}_{\mathcal{A}_n}(X_{m+1:n}-X_{m:n})\{f(\check{X})-f(X_{m:n})\}\big|\Big] \\
&\qquad \le C_1 E\big(\mathbf{I}_{\mathcal{A}_n}|X_{m+1:n}-X_{m:n}|^2\big)^{1/2} E\big(\mathbf{I}_{\mathcal{A}_n}|f(\check{X})-f(X_{m:n})|^2\big)^{1/2} \\
&\qquad \le C_1 E\big(\mathbf{I}_{\mathcal{A}_n}|(X_{m+1:n}-X_{m:n})|^2\big)^{1/2} E\big(\mathbf{I}_{\mathcal{A}_n}|X_{m+1:n}-X_{m:n}|^{2\nu}\big)^{1/2} \\
&\qquad \le C_1 E\big(\mathbf{I}_{\mathcal{A}_n}|X_{m+1:n}-X_{m:n}|^2\big)^{1/2} E\big(\mathbf{I}_{\mathcal{A}_n}|X_{m+1:n}-X_{m:n}|^3\big)^{\nu/3} \\
&\qquad \le C_1 O(n^{-2})^{1/2} O(n^{-3})^{\nu/3} \\
&\qquad \le C_1 O(n^{-1-\nu}) \\
&\qquad = o(n^{-1}),
\end{aligned} \tag{4.111}$$

where $0 < \nu \le 1$ and C_1 is a generic constant that may differ from line to line.

From (4.111), (4.109), by further Taylor expansion of

$$1 - \frac{1}{2}(U_{m+2:n}-U_{m:n}) f^{(1)}(\bar{X}) f(\bar{X})^{-3} f(X_{m:n})^{-1},$$

and since a version of (4.110) holds on $\mathcal{A}_n$ with

$$(n+1)\hat{\alpha} - m = \frac{m}{n} + \{\Delta + (n\alpha - m)(1-\Delta)\}$$

and $\alpha = \frac{m}{n} + O(n^{-1})$ (see Theorem 4.1), observe that

$$\begin{aligned}
P(X \in \widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha}) &= E\{F(\hat{\xi}_{\mathrm{Q2},\hat{\beta}})(\mathbf{I}_{\mathcal{A}_n} + \mathbf{I}_{\mathcal{A}_n^c})\} \\
&= E\Bigg[U_{m:n} + \alpha(U_{m+1:n}-U_{m:n}) + (U_{m+1:n}-U_{m:n})\Big\{\frac{U_{m+1:n}-U_{m:n}}{U_{m+2:n}-U_{m:n}} \\
&\qquad + (n\alpha - m)\frac{U_{m+2:n}-U_{m+1:n}}{U_{m+2:n}-U_{m:n}}\Big\}\Bigg] + o(n^{-1}) \\
&= \frac{m+\alpha}{n+1} + E\Bigg[(U_{m+1:n}-U_{m:n})\Big\{\frac{U_{m+1:n}-U_{m:n}}{U_{m+2:n}-U_{m:n}} \\
&\qquad + (n\alpha - m)\frac{U_{m+2:n}-U_{m+1:n}}{U_{m+2:n}-U_{m:n}}\Big\}\Bigg] + o(n^{-1}) \\
&= \frac{m+\alpha}{n+1} + \frac{n\alpha - m}{n+1} + (m+1-n\alpha)E\Big\{\frac{(U_{m+1:n}-U_{m:n})^2}{U_{m+2:n}-U_{m:n}}\Big\} + o(n^{-1}) \\
&= \alpha + (m+1-n\alpha)E\Big\{\frac{(U_{m+1:n}-U_{m:n})^2}{U_{m+2:n}-U_{m:n}}\Big\} + o(n^{-1})
\end{aligned}$$

Let $\mathcal{Z} = (Z_1, Z_2, \ldots, Z_n, Z_{n+1})^\top$ denote $n+1$ independent and identically distributed random variables drawn from a standard exponential population with $F_Z(z) = 1 - e^{-z}$ as the probability distribution function. From Reiss (1989, Theorem 1.6.7) it follows that

$$U_{m+1:n} - U_{m:n} \stackrel{d}{=} \frac{Z_{m+1}}{\sum_{j=1}^{n+1} Z_j} \quad \text{and} \quad U_{m+2:n} - U_{m:n} \stackrel{d}{=} \frac{Z_{m+2} + Z_{m+1}}{\sum_{j=1}^{n+1} Z_j}.$$

Therefore,

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha}) = \alpha + \frac{m+1-n\alpha}{n} E\left(\frac{Z_{m+1}^2}{Z_{m+1} + Z_{m+2}}\right) + o(n^{-1}). \tag{4.112}$$

Since $E\{Z_{m+1}^2/(Z_{m+1} + Z_{m+2})\} = 2/3$ and by using properties of lim sup (Royden, 1988) it can be observed from (4.112) that

$$\begin{aligned}
\limsup_{n\to\infty} n\left|P(X \in \widehat{\mathcal{I}}_{\mathrm{LC,Q2},\alpha}) - \alpha\right| &= \limsup_{n\to\infty} n\left|\frac{2(m+1-n\alpha)}{3n} + o(n^{-1})\right| \\
&\geq \limsup_{n\to\infty} \left\{\tfrac{2}{3}|(m+1-n\alpha)| - |o(1)|\right\} \\
&\geq \limsup_{n\to\infty} \tfrac{2}{3}|(m+1-n\alpha)| \\
&> \tfrac{2}{3}(\alpha + \eta) > 0,
\end{aligned}$$

where the sequence $\mathcal{S}$ depends on α and η. This concludes the proof. ∎

4.3.4 Statement and Proof of Theorem 4.4

Theorem 4.4 *Let the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables which are drawn from a population with F as probability distribution function and $f = F'$ as probability density function. Assume that the α-th population quantile $\xi_\alpha = F^{-1}(\alpha)$ is uniquely defined, that F has five bounded derivatives in a neighbourhood of ξ_α, and that $f(\xi_\alpha) > 0$. Furthermore, assume that the kernel K is a nonnegative, symmetric, and compactly supported probability density function with three bounded derivatives and that $E(|X|^l) < \infty$ for l sufficiently large. Let the nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{AC,Q2,\alpha}$ and $\widehat{\mathcal{I}}_{AC,Q3,\alpha}$ be defined at* (4.41). *Then*

$$P(X \in \widehat{\mathcal{I}}_{AC,Q2,\alpha}) = \alpha + O(n^{-2-2/3}) \qquad \text{if } h \sim cn^{-1/3} \tag{4.113}$$

and

$$P(X \in \widehat{\mathcal{I}}_{AC,Q3,\alpha}) = \alpha + O(n^{-3-2/5}) \qquad \text{if } h \sim cn^{-1/5} \tag{4.114}$$

where $c > 0$ is any positive constant and the predictand X is independent of the sample $\mathcal{X}$ but drawn from the same population.

Proof. Denote by $\mathcal{N}$ the neighbourhood of the α-th population quantile ξ_α in which the population distribution function F has five bounded derivatives, and select $\epsilon > 0$ sufficiently small such that

$$\mathcal{J} = \{\xi : |\xi - \xi_\alpha| \leq \epsilon\} \subset \mathcal{N}.$$

First concentrate on deriving the coverage error properties of the additive-calibrated prediction interval $\widehat{\mathcal{I}}_{\text{AC,Q3},\alpha}$ which has upper end-point given by $\hat{\xi}_{\text{Q3},\alpha} + \hat{a}_3$. Deriving the coverage error properties of the additive-calibrated prediction interval $\widehat{\mathcal{I}}_{\text{AC,Q2},\alpha}$ which has upper end-point given by $\hat{\xi}_{\text{Q2},\alpha} + \hat{a}_2$ is similar, and differences will be illustrated after the conclusion of result (4.114).

Using Hoeffding's inequality and Taylor expansion (see Theorem 4.1) it can be shown that

$$P(|X_{m-1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.115}$$

$$P(|X_{m:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.116}$$

$$P(|X_{m+1:n} - \xi_\alpha| > \epsilon) = O(n^{-\lambda}), \tag{4.117}$$

and

$$P(|\hat{\xi}_{\text{Q3},\alpha} + a_3 - \xi_\alpha| > \epsilon) = O(n^{-\lambda}) \tag{4.118}$$

for any $\lambda > 0$, whenever $|a_3| \leq n^{-\epsilon_1}$ for some $\epsilon_1 > 0$.

Let $\mathcal{A}_n = \mathcal{A}_{1,n} \cap \mathcal{A}_{2,n} \cap \mathcal{A}_{3,n} \cap \mathcal{A}_{4,n}$, where $\mathcal{A}_{1,n} = \{X_{m-1:n} \in \mathcal{N}\}$, $\mathcal{A}_{2,n} = \{X_{m:n} \in \mathcal{N}\}$, $\mathcal{A}_{3,n} = \{X_{m+1:n} \in \mathcal{N}\}$, and $\mathcal{A}_{4,n} = \{\hat{\xi}_{\text{Q3},\alpha} + a_2 \in \mathcal{N}\}$. Using Boole's inequality in conjunction with (4.115), (4.116), (4.117), and (4.118) it follows that

$$P(\mathcal{A}_n^c) = O(n^{-\lambda}) \tag{4.119}$$

for any $\lambda > 0$.

A consequence of (4.119) is that, without loss of generality, the population probability distribution function F can be assumed to be continuous with five bounded derivatives on the whole real line. Therefore, by the Inverse Function Theorem and Taylor expansion, it can be deduced that

$$\begin{aligned} X_{m-1+k:n} - X_{m-1:n} &= F^{-1}(U_{m-1:n} + U_{m-k+k:n} - U_{m-1:n}) - F^{-1}(U_{m-1:n}) \\ &= \sum_{r=1}^{3} \frac{1}{r!}\delta_k^r G^{(r)}(X_{m-1:n}) + \frac{1}{4!}\delta_k^4 G^{(4)}(\tilde{X}_k), \qquad k = 1, 2, \end{aligned} \tag{4.120}$$

where $f^{(j)}(x)$ for $j = 1, 2, 3, 4$ denotes the i-th derivative of f,

$$G^{(1)}(x) = f(x)^{-1},$$

$$G^{(2)}(x) = -f^{(1)}(x)f(x)^{-3},$$

$$G^{(3)}(x) = 3\{f^{(1)}(x)\}^2 f(x)^{-5} - f^{(2)}(x)f(x)^{-4},$$

$$G^{(4)}(x) = -f^{(3)}(x)f(x)^{-5} + 10f^{(1)}(x)f^{(2)}(x)f(x)^{-6} - 15\{f^{(1)}(x)\}^3 f(x)^{-7},$$

$$\delta_k = U_{m-1+k:n} - U_{m-1:n},$$

and $\tilde{X}_k$ lies on the line segment between $X_{m-1+k:n}$ and $X_{m-1:n}$.

Since $a + b + c = 1$ (see (4.15)) write

$$\begin{aligned}\hat{\xi}_{\mathrm{Q3},\alpha} &= aX_{m-1:n} + bX_{m:n} + cX_{m+1:n} \\ &= X_{m-1:n} + b(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}).\end{aligned}$$

Then, by Taylor expansion, observe that

$$\begin{aligned}F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3) &= F\{X_{m-1:n} + b(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}) + a_3\} \\ &= F(X_{m-1:n}) + \{b(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}) + a_3\}f(X_{m-1:n}) \\ &\quad + \frac{1}{2}\{b(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}) + a_3\}^2 f^{(1)}(X_{m-1:n}) \\ &\quad + \frac{1}{6}\{b(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}) + a_3\}^3 f^{(2)}(X_{m-1:n}) \\ &\quad + \frac{1}{24}\{a(X_{m:n} - X_{m-1:n}) + c(X_{m+1:n} - X_{m-1:n}) + a_3\}^4 f^{(3)}(\tilde{X}_1) \qquad (4.121)\end{aligned}$$

where $\tilde{X}_1$ lies on the line segment between $\hat{\xi}_{\mathrm{Q3},\alpha} + a_3$ and $X_{m-1:n}$. Substituting (4.120) into (4.121) it can be deduced that

$$\begin{aligned}F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3) &= U_{m-1:n} + b\delta_1 + c\delta_2 + \frac{1}{2}\{(b\delta_1 + c\delta_2)^2 - (b\delta_1^2 + c\delta_2^2)\}f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-2} \\ &\quad + \frac{1}{6}\big[(b\delta_1 + c\delta_2)^3 f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3} \\ &\qquad - 3(b\delta_1 + c\delta_2)(b\delta_1^2 + c\delta_2^2)f^{(1)}(X_{m-1:n})^2 f(X_{m-1:n})^{-4} \\ &\qquad + (b\delta_1^3 + c\delta_2^3)\{3f^{(1)}(X_{m-1:n})^2 f(X_{m-1:n})^{-4} \\ &\qquad - f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3}\}\big] \\ &\quad + a_3 f(X_{m-1:n}) + a_3(b\delta_1 + c\delta_2)f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-1} + R_n, \qquad (4.122)\end{aligned}$$

where R_n contains all the remaining terms from the expansion.

Since

$$P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}) = E\{F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3)\mathbf{I}_{\mathcal{A}_n} + F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3)\mathbf{I}_{\mathcal{A}_n^c}\},$$

$E\{F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3)\mathbf{I}_{\mathcal{A}_n^c}\} \leq P(\mathcal{A}_n^c) = O(n^{-\lambda}) \leq O(n^{-4} + a_3^2)$ (see (4.119)) whenever $\lambda \geq 4$, and a version of the expansion indexed by (4.122) holds on $\mathcal{A}_n$,

$$\begin{aligned}
P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}) &= E\big\{(U_{m-1:n} + b\delta_1 + c\delta_2)\mathbf{I}_{\mathcal{A}_n}\big\} && (4.123)\\
&\quad + E\big[\tfrac{1}{2}\{(b\delta_1 + c\delta_2)^2 - (b\delta_1^2 + c\delta_2^2)\}f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-2}\mathbf{I}_{\mathcal{A}_n}\big] && (4.124)\\
&\quad + E\Big(\tfrac{1}{6}\big[(b\delta_1 + c\delta_2)^3 f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3}\\
&\qquad - 3(b\delta_1 + c\delta_2)(b\delta_1^2 + c\delta_2^2)f^{(1)}(X_{m-1:n})^2 f(X_{m-1:n})^{-4}\\
&\qquad + (b\delta_1^3 + c\delta_2^3)\{3f^{(1)}(X_{m-1:n})^2 f(X_{m-1:n})^{-4}\\
&\qquad - f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3}\}\big]\mathbf{I}_{\mathcal{A}_n}\Big) && (4.125)\\
&\quad + E\Big[\{a_3 f(X_{m-1:n}) + a_3(b\delta_1 + c\delta_2)f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-1}\}\mathbf{I}_{\mathcal{A}_n}\Big] && (4.126)\\
&\quad + E(R_n\mathbf{I}_{\mathcal{A}_n}) + O(n^{-4} + a_3^2). && (4.127)
\end{aligned}$$

Using the fact that $U_{m-1:n}$, δ_1, δ_2, b, and c are bounded in n, and applying (4.119) for a suitable selection of $\lambda > 0$, it can be shown that (4.123) can be replaced by

$$\begin{aligned}
&E\big\{(U_{m-1:n} + b\delta_1 + c\delta_2)\big\}\\
&\qquad - E\big\{(U_{m-1:n} + b\delta_1 + c\delta_2)\mathbf{I}_{\mathcal{A}_n^c}\big\} = \frac{m-1}{n+1} + \frac{b}{n+1} + \frac{2c}{n+1} + O(n^{-4} + a_3^2).
\end{aligned}$$

From David (1981, Theorem 2.7) or Reiss (1989, Theorem 1.8.1) it follows that, conditional on $X_{m-1:n}$, the random vector $(\delta_1/\{1 - F(X_{m-1:n})\}, \delta_2/\{1 - F(X_{m-1:n})\})^\top$ has the same probability distribution function as $(U_{1:n-m+1}, U_{2:n-m+1})^\top$. Denote $U_{k:n-m+1}$ by V_k for $k = 1, 2$ which may be taken to be independent of $X_{m-1:n}$. Therefore it can be deduced that (4.124) can be replaced by

$$\begin{aligned}
&E\Big[\tfrac{1}{2}\{(bV_1 + cV_2)^2 - (bV_1^2 + cV_2^2)\}f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-2}\{1 - F(X_{m-1:n})\}^2\mathbf{I}_{\mathcal{A}_n}\Big]\\
&\qquad = \tfrac{1}{2}E\big[f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-2}\{1 - F(X_{m-1:n})\}^2\mathbf{I}_{\mathcal{A}_{1,n}}\big]\\
&\qquad\qquad E\big\{(bV_1 + cV_2)^2 - (bV_1^2 + cV_2^2)\big\}\mathbf{I}_{\mathcal{A}_n}\big\}. \qquad (4.128)
\end{aligned}$$

Noting that V_1, V_2, b, and c are bounded in n, and applying (4.119), it can be deduced that

$$\begin{aligned}
&E\big[\big\{(bV_1 + cV_2)^2 - (bV_1^2 + cV_2^2)\big\}\mathbf{I}_{\mathcal{A}_n}\big]\\
&\qquad = E\big\{(bV_1 + cV_2)^2 - (bV_1^2 + cV_2^2)\big\} + O(n^{-4} + a_3^2) \qquad (4.129)
\end{aligned}$$

for a suitable selection of $\lambda > 0$. Furthermore, since

$$E(V_1^2) = 2/\{(n-m+2)(n-m+3)\},$$
$$E(V_1V_2) = 3/\{(n-m+2)(n-m+3)\},$$

and

$$E(V_2^2) = 6/\{(n-m+2)(n-m+3)\},$$

it can be shown that (4.129) can be replaced by

$$E\Big[\{(bV_1+cV_2)^2-(bV_1^2+cV_2^2)\}\mathbf{I}_{\mathcal{A}_n}\Big] = 2(b^2-b)+6(c^2-c)+6bc+O(n^{-4}+a_3^2)$$
$$= O(n^{-4}+a_3^2), \tag{4.130}$$

because $2(b^2-b)+6(c^2-c)+6bc = (a-c)^2+a^2+c^2-2(a+c) = 0$ by the selection of the bounded coefficients in Theorem 4.1 (see Subsection 4.1.3). Therefore, substituting (4.130) into (4.128), it follows that

$$E\Big[\tfrac{1}{2}\{(bV_1+cV_2)^2-(bV_1^2+cV_2^2)\}f^{(1)}(X_{m-1:n})f(X_{m-1:n})^{-2}\{1-F(X_{m-1:n})\}^2\mathbf{I}_{\mathcal{A}_n}\Big]$$
$$= O(n^{-4}+a_3^2). \tag{4.131}$$

From technical arguments analogous to those used when determining (4.131), it can be deduced that

$$E\Big(\tfrac{1}{6}\big[(b\delta_1+c\delta_2)^3f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3}$$
$$-3(b\delta_1+c\delta_2)(b\delta_1^2+c\delta_2^2)f^{(1)}(X_{m-1:n})^2f(X_{m-1:n})^{-4}$$
$$+(b\delta_1^3+c\delta_2^3)\{3f^{(1)}(X_{m-1:n})^2f(X_{m-1:n})^{-4}$$
$$-f^{(2)}(X_{m-1:n})f(X_{m-1:n})^{-3}\}\big]\mathbf{I}_{\mathcal{A}_n}\Big) = \tfrac{1}{6}\Big[\gamma_3(n)E\{(bV_1+cV_2)^3-$$
$$3\gamma_2(n)E\{(bV_1+cV_2)(bV_1^2+cV_2^2)\}$$
$$+\{3\gamma_2(n)-\gamma_3(n)\}E(bV_1^3+cV_2^3)\Big]+O(n^{-4}+a_3^2), \tag{4.132}$$

where

$$\gamma_2(n) = E\Big[\frac{f^{(1)}(X_{m-1:n})^2}{f(X_{m-1:n})^4}\{1-F(X_{m-1:n})\}^3\mathbf{I}_{\mathcal{A}_{1,n}}\Big] \tag{4.133}$$

and

$$\gamma_3(n) = E\Big[\frac{f^{(2)}(X_{m-1:n})}{f(X_{m-1:n})^3}\{1 - F(X_{m-1:n})\}^3 \mathbf{I}_{\mathcal{A}_{1,n}}\Big]. \tag{4.134}$$

Since a and c are selected such that $c - a = (n+1)\alpha - m$ (see Theorem 4.1), it can be shown that

$$\begin{aligned} E(U_{m-1:n} + b\delta_1 + c\delta_2) &= \frac{m - 1 + b + 2c}{n+1} \\ &= \frac{m - a + c}{n+1} = \alpha. \end{aligned} \tag{4.135}$$

Properties of the coefficients a, b, c, and the Taylor expansion

$$f(X_{m-1:n}) = f\{F^{-1}(\alpha + U_{m-1:n} - \alpha)\} = f(\xi_\alpha) + (U_{m-1:n} - \alpha) f^{(1)}(\tilde{X}_3) f(\tilde{X}_3)^{-1},$$

where $\tilde{X}_3$ lies on the line segment between ξ_α and $X_{m-1:n}$, can be used to show that

$$\begin{aligned} E\Big[\{a_3 f(X_{m-1:n}) + a_3(b\delta_1 + c\delta_2) f^{(1)}(X_{m-1:n}) f(X_{m-1:n})^{-1}\}\mathbf{I}_{\mathcal{A}_n}\Big] \\ = a_3 f(\xi_\alpha) + O(n^{-4} + a_3^2). \end{aligned} \tag{4.136}$$

Noting that $n^3 E\{(bV_1 + cV_2)^3\}$, $n^3 E\{(bV_1 + cV_2)(bV_1^2 + cV_2^2)\}$, and $n^3 E(bV_1^3 + cV_2^3)$ are bounded in n, and using Young's inequality, it can be shown that (4.127) can be replaced by $E(R_n \mathbf{I}_{\mathcal{A}_n}) = O(n^{-4} + a_3^2)$. Further replacing (4.123), (4.124), (4.125), and (4.126) by (4.135), (4.131), (4.132), and (4.136), respectively, observe that

$$P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}) = \alpha + a_3 f(\xi_\alpha) + n^{-3}\{c_2(n)\gamma_2(n) + c_3(n)\gamma_3(n)\} + O(n^{-4} + a_3^2), \tag{4.137}$$

where $c_2(n)$ and $c_3(n)$ are bounded and do not depend on the population distribution function, and $\gamma_2(n)$ and $\gamma_3(n)$ are defined at (4.133) and (4.134), respectively.

The smoothed bootstrap version of (4.137) replaces $P(X \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}) = E\{F(\hat{\xi}_{\mathrm{Q3},\alpha} + a_3)\}$ by $P(X^\dagger \in \mathcal{I}^\dagger_{\mathrm{AC,Q3},\alpha} \,|\, \mathcal{X}) = E\{\hat{F}(\hat{\xi}^\dagger_{\mathrm{Q3},\alpha} + a_3) \,|\, \mathcal{X}\}$, where the same notation as Subsection 4.2.2 is employed. To calculate such an expansion, begin by defining subsets $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_9$ of $\mathcal{C}$, the class of all possible samples $\mathcal{X}$.

Rosenthal's inequality (Burkholder, 1973; Hall and Heyde, 1980, p. 23) states that if $S_0 = 0$ and $\{S_i, \mathcal{F}_i, 1 \le i \le n\}$ is a martingale, then for $2 \le p < \infty$ there exist constants C_1 and C_2

depending only on p such that

$$C_1\left(E\Big[\sum_{i=1}^{n} E\{(S_i - S_{i-1})^2 \mid \mathcal{F}_{i-1}\}\Big]^{p/2} + \sum_{i=1}^{n} E|S_i - S_{i-1}|^p\right)$$
$$\leq E|S_n|^p \leq C_2\left(E\Big[\sum_{i=1}^{n} E\{(S_i - S_{i-1})^2 \mid \mathcal{F}_{i-1}\}\Big]^{p/2} + \sum_{i=1}^{n} E|S_i - S_{i-1}|^p\right). \quad (4.138)$$

Given the kernel density estimator $\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K\big(\frac{x-X_i}{h}\big)$ and for fixed $x \in \mathcal{J}$, set $S_n = \hat{f}(x) - E\{\hat{f}(x)\}$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. Then $\{S_i, \mathcal{F}_i, 1 \leq i \leq n\}$ is a martingale, indeed any sum of independent random variables is a martingale, and (4.138) specialises to give

$$E\Big|\hat{f}(x) - E\{\hat{f}(x)\}\Big|^{2k} \leq C_2\Bigg[E\Big\{\sum_{i=1}^{n} E\Big((nh)^{-2}\Big[K\big(\tfrac{x-X_i}{h}\big) - E\{K\big(\tfrac{x-X_i}{h}\big)\}\Big]^2\Big|\mathcal{F}_{i-1}\Big)\Big\}^k$$
$$+ \sum_{i=1}^{n} E\Big|(nh)^{-1}\Big[K\big(\tfrac{x-X_i}{h}\big) - E\{K\big(\tfrac{x-X_i}{h}\big)\}\Big]\Big|^{2k}\Bigg]$$
$$\leq C_2\Bigg[n^{-k}h^{-2k}\Big\{E\Big(\Big[K\big(\tfrac{x-X}{h}\big) - E\{K\big(\tfrac{x-X}{h}\big)\}\Big]^2\Big)\Big\}^k$$
$$+ n^{-2k+1}h^{-2k}E\Big|K\big(\tfrac{x-X}{h}\big) - E\{K\big(\tfrac{x-X}{h}\big)\}\Big|^{2k}\Bigg], \quad (4.139)$$

because X_i for $i = 1, 2, \ldots, n$ are independent.

The property

$$E\Big(\Big[K\big(\tfrac{x-X}{h}\big) - E\{K\big(\tfrac{x-X}{h}\big)\}\Big]^2\Big) \leq E\Big[\Big\{K\big(\tfrac{x-X}{h}\big)\Big\}^2\Big]$$
$$= \int \Big\{K\big(\tfrac{x-y}{h}\big)\Big\}^2 f(y)\,dy$$
$$= h\int \{K(t)\}^2 f(x-ht)\,dt = O(h), \quad (4.140)$$

the c_r-inequality (Prakasa Rao, 1987, Proposition 1.14.12), can be used to shown that

$$E\Big|K\big(\tfrac{x-X}{h}\big) - E\{K\big(\tfrac{x-X}{h}\big)\}\Big|^{2k} \leq C\Big(E\Big[\{K\big(\tfrac{x-X}{h}\big)\}^{2k}\Big] + \Big[E\{K\big(\tfrac{x-X}{h}\big)\}\Big]^{2k}\Big)$$
$$= O(h). \quad (4.141)$$

Therefore, combining (4.140) and (4.141) with (4.139) it follows that

$$E\Big|\hat{f}(x) - E\{\hat{f}(x)\}\Big|^{2k} = O\{(nh)^{-k}\} \quad (4.142)$$

uniformly in $x \in \mathcal{J}$.

Let $K^{(j)}$, for $j = 1, 2, 3$, denote the j-th derivative of the kernel K and let

$$\hat{f}^{(1)}(x) = (nh^2)^{-1} \sum_{i=1}^{n} K^{(1)}\left(\tfrac{x-X_i}{h}\right),$$
$$\hat{f}^{(2)}(x) = (nh^3)^{-1} \sum_{i=1}^{n} K^{(2)}\left(\tfrac{x-X_i}{h}\right),$$

and

$$\hat{f}^{(3)}(x) = (nh^4)^{-1} \sum_{i=1}^{n} K^{(3)}\left(\tfrac{x-X_i}{h}\right)$$

denote the first, second, and third derivatives of the kernel density estimator $\hat{f}(x)$, respectively. Since $\hat{f}^{(1)}(x) - E\{\hat{f}^{(1)}(x)\}$ is a martingale and $K^{(1)}$ is bounded, (4.138) specialises to give

$$\begin{aligned} E\left|\hat{f}^{(1)}(x) - E\{\hat{f}^{(1)}(x)\}\right|^{2k} \leq C_2 \Bigg[& n^{-k}h^{-4k}\left\{E\left(\left[K^{(1)}\left(\tfrac{x-X}{h}\right) - E\{K^{(1)}\left(\tfrac{x-X}{h}\right)\}\right]^2\right)\right\}^k \\ & + n^{-2k+1}h^{-4k}E\left|K^{(1)}\left(\tfrac{x-X}{h}\right) - E\{K^{(1)}\left(\tfrac{x-X}{h}\right)\}\right|^{2k}\Bigg] \\ = O\{(nh^3)^{-k}\} & \end{aligned} \tag{4.143}$$

uniformly in $x \in \mathcal{J}$. Additionally it can be shown that

$$E\left|\hat{F}(x) - E\{\hat{F}(x)\}\right|^{2k} = O\{(nh^{-1})^{-k}\}, \tag{4.144}$$
$$E\left|\hat{f}^{(2)}(x) - E\{\hat{f}^{(2)}(x)\}\right|^{2k} = O\{(nh^5)^{-k}\}, \tag{4.145}$$

and

$$E\left|\hat{f}^{(3)}(x) - E\{\hat{f}^{(3)}(x)\}\right|^{2k} = O\{(nh^7)^{-k}\} \tag{4.146}$$

uniformly in $x \in \mathcal{J}$.

By Markov's inequality, (4.142), and for k sufficiently large

$$\begin{aligned} P\left[\left|\hat{f}(x) - E\{\hat{f}(x)\}\right| > (nh)^{-1/2}n^{\gamma}\right] &\leq \frac{E\left|\hat{f}(x) - E\{\hat{f}(x)\}\right|^{2k}}{(nh)^{-k}n^{2k\gamma}} \\ &= O(n^{-2k\gamma}) \\ &= O(n^{-\lambda_1}) \end{aligned} \tag{4.147}$$

for any $\lambda_1, \gamma > 0$, uniformly in $x \in \mathcal{J}$.

Let $\mathcal{S}_n$ denote any collection of no more than $O(n^{\lambda_2})$ elements of the range $\mathcal{J}$, for an arbitrarily large but fixed value of $\lambda_2 > 0$. From (4.147) it follows that

$$\sup_{x \in \mathcal{S}_n} P\left[\left|\hat{f}(x) - E\{\hat{f}(x)\}\right| > (nh)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda_1}).$$

Hence, taking $\lambda_1 > \lambda_2 + \lambda$ and using Boole's inequality, it follows that

$$\begin{aligned} P\left[\sup_{x \in \mathcal{S}_n} \left|\hat{f}(x) - E\{\hat{f}(x)\}\right| > (nh)^{-1/2} n^{\gamma}\right] &\leq \sum_{x \in \mathcal{S}_n} P\left[\left|\hat{f}(x) - E\{\hat{f}(x)\}\right| > (nh)^{-1/2} n^{\gamma}\right] \\ &= O(n^{\lambda_2}) O(n^{-\lambda_1}) = O(n^{-\lambda}) \end{aligned}$$

for any $\lambda > 0$.

Additionally it can be shown that

$$P\left[\sup_{x \in \mathcal{S}_n} \left|\hat{f}^{(1)}(x) - E\{\hat{f}^{(1)}(x)\}\right| > (nh^3)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}),$$
$$P\left[\sup_{x \in \mathcal{S}_n} \left|\hat{f}^{(2)}(x) - E\{\hat{f}^{(2)}(x)\}\right| > (nh^5)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}),$$

and

$$P\left[\sup_{x \in \mathcal{S}_n} \left|\hat{f}^{(3)}(x) - E\{\hat{f}^{(3)}(x)\}\right| > (nh^7)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda})$$

for any $\lambda, \gamma > 0$.

Since K is Hölder continuous and by using Young's form of Taylor theorem it can be deduced that $|K(x_1) - K(x_2)|$ and $|K^{(j)}(x_1) - K^{(j)}(x_2)|$ for $j = 1, 2, 3$ are bounded for $x_1, x_2 \in \mathcal{J}$. Therefore, it can be shown that

$$P\left[\sup_{x \in \mathcal{J}} \left|\hat{f}(x) - E\{\hat{f}(x)\}\right| > (nh)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}), \tag{4.148}$$
$$P\left[\sup_{x \in \mathcal{J}} \left|\hat{f}^{(1)}(x) - E\{\hat{f}^{(1)}(x)\}\right| > (nh^3)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}), \tag{4.149}$$
$$P\left[\sup_{x \in \mathcal{J}} \left|\hat{f}^{(2)}(x) - E\{\hat{f}^{(2)}(x)\}\right| > (nh^5)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}), \tag{4.150}$$

and

$$P\left[\sup_{x \in \mathcal{J}} \left|\hat{f}^{(3)}(x) - E\{\hat{f}^{(3)}(x)\}\right| > (nh^7)^{-1/2} n^{\gamma}\right] = O(n^{-\lambda}) \tag{4.151}$$

for any $\lambda, \gamma > 0$. Similar deductive reasoning was used, for example, by Bowman, Hall, and Prvan (1998).

By substitution and Taylor expansion, observe that

$$\begin{aligned} E\{\hat{f}(x)\} &= h^{-1}\int K\left(\tfrac{x-y}{h}\right) f(y)\,dy \\ &= \int K(t) f(x-ht)\,dt \\ &= \int K(t)\{f(x) - htf^{(1)}(x) + (ht)^2 f^{(2)}(x-\nu_1 ht)\}\,dt \\ &= f(x) + O(h^{-2}) \end{aligned} \tag{4.152}$$

uniformly in $x \in \mathcal{J}$, where $0 < \nu_1 < 1$.

Using the fact that the kernel K has compact support and employing substitution, integration by parts, and Taylor expansion, it can be shown that

$$\begin{aligned} E\{\hat{f}^{(1)}(x)\} &= h^{-2}\int K^{(1)}\left(\tfrac{x-y}{h}\right) f(y)\,dy \\ &= h^{-1}\int K^{(1)}(t) f(x-ht)\,dt \\ &= h^{-1}\{K(t)f(x-ht)\,\big| + h\int K(t) f^{(1)}(x-ht)\,dt\} \\ &= \int K(t) f^{(1)}(x-ht)\,dt \\ &= \int K(t)\{f^{(1)}(x) - htf^{(2)}(x) + (ht)^2 f^{(3)}(x-\nu_2 ht)\}\,dt \\ &= f^{(1)}(x) + O(h^2) \end{aligned} \tag{4.153}$$

uniformly in $x \in \mathcal{J}$, where $0 < \nu_2 < 1$.

Furthermore,

$$E\{\hat{f}^{(2)}(x)\} = f^{(2)}(x) + O(h^2)$$

and

$$E\{\hat{f}^{(3)}(x)\} = f^{(3)}(x) + O(h^2) \tag{4.154}$$

uniformly in $x \in \mathcal{J}$, where (4.154) is derived using Young's form of Taylor's Theorem.

Assume that $C_1 n^{-1+\delta} \leq h \leq C_2 n^{-\delta}$, where $\delta \in (0, \frac{1}{2})$ and C_1 and C_2 denote positive constants.

Taking $0 < \gamma < \delta/2$ and by using (4.148) and (4.152), it can be shown that

$$\begin{aligned} P\Big\{ \sup_{x\in\mathcal{J}} |\hat{f}(x) - f(x)| \geq \gamma_1 \Big\} &\leq P\Big[\sup_{x\in\mathcal{J}} |\hat{f}(x) - E\{\hat{f}(x)\} + E\{\hat{f}(x)\} - f(x)| \geq (nh)^{-1/2} n^{\gamma} \Big] \\ &\leq P\Big[\sup_{x\in\mathcal{J}} |\hat{f}(x) - E\{\hat{f}(x)\}| \geq (nh)^{-1/2} n^{\gamma} \Big] \\ &\quad + P\Big[\sup_{x\in\mathcal{J}} |E\{\hat{f}(x)\} - f(x)| \geq (nh)^{-1/2} n^{\gamma} \Big] = O(n^{-\lambda}), \end{aligned}$$

for any γ_1, $\lambda > 0$. Therefore,

$$\begin{aligned} P\Big\{ \hat{f}(y) \leq \inf_{x\in\mathcal{J}} f(x) - \gamma_1, \text{ for some } y \in \mathcal{J} \Big\} &\leq P\Big\{ \sup_{x\in\mathcal{J}} |\hat{f}(x) - f(x)| \geq \gamma_1 \Big\} \\ &= O(n^{-\lambda}) \end{aligned}$$

and

$$\begin{aligned} P\Big\{ \inf_{x\in\mathcal{J}} \hat{f}(x) > \inf_{x\in\mathcal{J}} f(x) - \gamma_1 \Big\} &\leq P\Big\{ \hat{f}(y) > \inf_{x\in\mathcal{J}} f(x) - \gamma_1, \text{ for some } y \in \mathcal{J} \Big\} \\ &= 1 - P\Big\{ \hat{f}(y) \leq \inf_{x\in\mathcal{J}} f(x) - \gamma_1, \text{ for some } y \in \mathcal{J} \Big\} \\ &= 1 - O(n^{-\lambda}). \end{aligned}$$

Hence,

$$\begin{aligned} P\Big\{ \inf_{x\in\mathcal{J}} \hat{f}(x) \leq \inf_{x\in\mathcal{J}} f(x) - \gamma_1 \Big\} &= 1 - P\Big\{ \inf_{x\in\mathcal{J}} \hat{f}(x) > \inf_{x\in\mathcal{J}} f(x) - \gamma_1 \Big\} \\ &= O(n^{-\lambda}) \end{aligned} \tag{4.155}$$

for any $\lambda > 0$.

From (4.155) it follows that for some constant $C_1 > 0$, the set

$$\mathcal{E}_1 = \Big\{ \inf_{x\in\mathcal{J}} \hat{f}(x) > C_1 \Big\}$$

satisfies $P(\mathcal{E}_1^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Using the fact that $|\hat{f}(x)| - |E\{\hat{f}(x)\}| \leq |\hat{f}(x) - E\{\hat{f}(x)\}|$, and employing the probability bound given by (4.148), it can be shown that

$$\begin{aligned} &P\Big[|\hat{f}(y)| > \sup_{x\in\mathcal{J}} |E\{\hat{f}(x)\}| + (nh)^{-1/2} n^{\gamma}, \text{ for some } y \in \mathcal{J} \Big] \\ &\qquad \leq P\big[|\hat{f}(y)| > |E\{\hat{f}(y)\}| + (nh)^{-1/2} n^{\gamma}, \text{ for some } y \in \mathcal{J} \big] \\ &\qquad \leq P\big[|\hat{f}(y)| - |E\{\hat{f}(y)\}| > (nh)^{-1/2} n^{\gamma}, \text{ for some } y \in \mathcal{J} \big] \\ &\qquad \leq P\Big[\sup_{x\in\mathcal{J}} \big| \hat{f}(x) - E\{\hat{f}(x)\} \big| > (nh)^{-1/2} n^{\gamma} \Big] \\ &\qquad = O(n^{-\lambda}). \end{aligned}$$

Therefore,

$$P\{\sup_{x\in\mathcal{J}}|\hat{f}(x)| \le (nh)^{-1/2}n^{\gamma}\} \le P\Big[\sup_{x\in\mathcal{J}}|\hat{f}(x)| \le \sup_{x\in\mathcal{J}}|E\{\hat{f}(x)\}| + (nh)^{-1/2}n^{\gamma}\Big]$$
$$\le P\big[|\hat{f}(y)| \le \sup_{x\in\mathcal{J}}|E\{\hat{f}(x)\}| + (nh)^{-1/2}n^{\gamma}, \text{ for some } y \in \mathcal{J}\big]$$
$$= 1 - O(n^{-\lambda}) \tag{4.156}$$

for any λ, $\gamma > 0$. Hence, for any $\gamma > 0$ the set

$$\mathcal{E}_2 = \left\{\sup_{x\in\mathcal{J}}|\hat{f}(x)| \le (nh)^{-1/2}n^{\gamma}\right\}$$

satisfies $P(\mathcal{E}_2^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Using the fact that $|\hat{f}^{(j)}(x)| - |E\{\hat{f}^{(j)}(x)\}| \le |\hat{f}^{(j)}(x) - E\{\hat{f}^{(j)}(x)\}|$ for $j = 1, 2, 3$ and employing the probability bounds (4.149), (4.150), and (4.151) for any $\gamma > 0$, it can be shown that the sets

$$\mathcal{E}_3 = \left\{\sup_{x\in\mathcal{J}}|\hat{f}^{(1)}(x)| \le (nh^3)^{-1/2}n^{\gamma}\right\},$$
$$\mathcal{E}_4 = \left\{\sup_{x\in\mathcal{J}}|\hat{f}^{(2)}(x)| \le (nh^5)^{-1/2}n^{\gamma}\right\},$$

and

$$\mathcal{E}_5 = \left\{\sup_{x\in\mathcal{J}}|\hat{f}^{(3)}(x)| \le (nh^7)^{-1/2}n^{\gamma}\right\}$$

satisfy $P(\mathcal{E}_3^c) = O(n^{-\lambda})$, $P(\mathcal{E}_4^c) = O(n^{-\lambda})$, and $P(\mathcal{E}_5^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Using integration by parts and arguments in Falk and Reiss (1989), observe that the kernel distribution function $\hat{F}$ satisfies

$$\begin{aligned}
\hat{F}(t) &= \int_{-\infty}^{t}\hat{f}(s)\,ds \\
&= \int_{-\infty}^{t}\sum_{i=1}^{n}\frac{1}{nh}K\Big(\frac{s-X_i}{h}\Big)\,ds \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{1}{h}K\Big(\frac{s-X_i}{h}\Big)\,ds \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{t-X_i}{h}}K(z)\,dz \\
&= \int\int_{-\infty}^{\frac{t-x}{h}}K(z)\,dz\,F_n(dx) \\
&= \int\frac{1}{h}K\Big(\frac{t-x}{h}\Big)F_n(x)\,dx \\
&= \int K(x)F_n(t-xh)\,dx,
\end{aligned} \tag{4.157}$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{(-\infty,x]}(X_i)$ denotes the empirical distribution function of the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$. From Taylor expansion, and (4.157), it can be shown that

$$E\{\hat{F}(t)\} = \int K(x) E\{F_n(t - xh)\}\, dx = F(t) + O(h^2). \tag{4.158}$$

Denote the α-th quantile of the kernel distribution function $\hat{F}$ by

$$\hat{\tau}_\alpha = \hat{F}^{-1}(\alpha) = \inf\{\hat{\tau} : \hat{F}(\hat{\tau}) \geq \alpha\}$$

and observe that

$$P(\hat{\tau}_\alpha \notin \mathcal{J}) = P(\hat{\tau}_\alpha > \xi_\alpha + \epsilon) + P(\hat{\tau}_\alpha < \xi_\alpha - \epsilon). \tag{4.159}$$

Consider the first term on the right hand side of (4.159). Then, using Hoeffding's inequality, properties of the kernel distribution function and (4.157) in conjunction with the compact support of the kernel K denoted by $[-C_{\mathrm{K}}, C_{\mathrm{K}}]$ for the positive constant C_{K}, it follows that

$$\begin{aligned}
P(\hat{\tau}_\alpha > \xi_\alpha + \epsilon) &= P\{\alpha > \hat{F}(\xi_\alpha + \epsilon)\} \\
&= P\{\alpha \int K(x)\, dx > \hat{F}(\xi_\alpha + \epsilon)\} \\
&= P\Big[\int K(x)\{F_n(\xi_\alpha + \epsilon - hx) - \alpha\}\, dx < 0\Big] \\
&\leq P\{F_n(\xi_\alpha + \epsilon - hC_{\mathrm{K}}) - F(\xi_\alpha) < 0\} \\
&= P\Big[\sum_{i=1}^{n} \{\hat{V}_i - E(\hat{V}_i)\} > n\{F(\xi_\alpha + \epsilon - hC_{\mathrm{K}}) - \alpha\}\Big] \\
&\leq e^{-2nd_1^2},
\end{aligned} \tag{4.160}$$

where $\hat{V}_i = \mathbf{I}_{(\xi_\alpha + \epsilon - hC_{\mathrm{K}}, \infty)}(X_i)$ and $d_1 = F(\xi_\alpha + \epsilon - hC_{\mathrm{K}}) - \alpha$.

Consider the second term on the right hand side of (4.159). Analogous to (4.160),

$$\begin{aligned}
P(\hat{\tau}_\alpha < \xi_\alpha - \epsilon) &\leq P\Big[\int K(x)\{\alpha - F_n(\xi_\alpha - \epsilon - hx)\} \leq 0\Big] \\
&= P\{\alpha - F_n(\xi_\alpha - \epsilon + C_{\mathrm{K}}h) \leq 0\} \leq e^{-2nd_2^2},
\end{aligned} \tag{4.161}$$

where $d_2 = \alpha - F(\xi_\alpha - \epsilon + C_{\mathrm{K}}h)$.

Substituting bounds (4.160) and (4.161) into (4.159) it follows that

$$P(\hat{\tau}_\alpha \notin \mathcal{J}) = O(n^{-\lambda})$$

for any $\lambda > 0$, because $\min(d_1, d_2) \geq \{\lambda/2\log(n)/n\}^{1/2}$ for n sufficiently large. Hence the set

$$\mathcal{E}_6 = \{\hat{\tau}_\alpha \in \mathcal{J}\}$$

satisfies $P(\mathcal{E}_6^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Let the smoothed bootstrap sample $\mathcal{X}^\dagger = (X_1^\dagger, X_2^\dagger, \ldots, X_n^\dagger)^\top$ denote, conditional on $\mathcal{X}$, n independent and identically distributed random variables which are drawn from a population with $\hat{F}$ as probability distribution function and let $F_n^\dagger(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{I}_{(-\infty,x]}(X_i^\dagger)$ denote the empirical distribution function of the smoothed bootstrap sample $\mathcal{X}^\dagger$. Let $X_{1:n}^\dagger < X_{2:n}^\dagger < \cdots < X_{n:n}^\dagger$, with $X_{i:n}^\dagger$ denoting the i-th order statistic of the smoothed bootstrap sample $\mathcal{X}^\dagger$. Then observe that

$$\begin{aligned} P(X_{m:n}^\dagger > \xi_\alpha + \epsilon) &= P(F_n^\dagger(\xi_\alpha + \epsilon) < \frac{m}{n}) \\ &\leq P\{1 - F_n^\dagger(\xi_\alpha + \epsilon) > 1 - \frac{m}{n}\} \\ &\leq P\{|\hat{F}(\xi_\alpha + \epsilon) - F_n^\dagger(\xi_\alpha + \epsilon)| > \frac{1}{2}(1 - \frac{m}{n})\} \\ &\qquad + P\{1 - \hat{F}(\xi_\alpha + \epsilon) > \frac{1}{2}(1 - \frac{m}{n})\} \qquad (4.162) \\ &= O(n^{-\lambda}), \qquad (4.163) \end{aligned}$$

where m satisfies (4.16). Both terms on the right hand side of (4.162) equal $O(n^{-\lambda})$ for any $\lambda > 0$. Similarly, it can be shown that,

$$P(X_{m:n}^\dagger < \xi_\alpha - \epsilon) = O(n^{-\lambda}) \qquad (4.164)$$

for any $\lambda > 0$.

Employing the bounds given by (4.163) and (4.164) it follows that

$$\begin{aligned} P(X_{m:n}^\dagger \notin \mathcal{J}) &= P(X_{m:n}^\dagger > \xi_\alpha + \epsilon) + P(X_{m:n}^\dagger < \xi_\alpha - \epsilon) \\ &= O(n^{-\lambda}) \end{aligned}$$

for any $\lambda > 0$. Therefore the set $\mathcal{E}_7 = \{X_{m:n}^\dagger \in \mathcal{J}\}$ satisfies $P(\mathcal{E}_7^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Analogously, it can be shown that the sets $\mathcal{E}_8 = \{X_{m-1:n}^\dagger \in \mathcal{J}\}$ and $\mathcal{E}_9 = \{X_{m+1:n}^\dagger \in \mathcal{J}\}$ satisfy $P(\mathcal{E}_8^c) = O(n^{-\lambda})$ and $P(\mathcal{E}_9^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

Throughout the remainder of this proof set $\mathcal{E} = \cap_{i=1}^9 \mathcal{E}_i$. Then, by Boole's inequality, $P(\mathcal{E}^c) = O(n^{-\lambda})$ for any $\lambda > 0$.

The construction of the smoothed bootstrap expansion for $P(X^\dagger \in \mathcal{I}_{\mathrm{AC,Q3},\alpha}^\dagger \,|\, \mathcal{X})$ proceeds analogously to the construction of (4.137) except that $\mathcal{A}_n$ is replaced by $\mathcal{E}$, F is replaced by $\hat{F}$,

and the unconditional expectation is replaced by the expectation conditional on the sample $\mathcal{X}$. Therefore, since $h \sim cn^{-1/5}$, by arguing on $\mathcal{E}$, and from the Borel-Cantelli lemma,

$$P(X^\dagger \in \mathcal{I}^\dagger_{\mathrm{AC,Q3},\alpha} \,|\, \mathcal{X}) = E\{\hat{F}(\hat{\xi}^\dagger_{\mathrm{Q3},\alpha} + a_3)(\mathbf{I}_\mathcal{E} + \mathbf{I}_{\mathcal{E}^c}) \,|\, \mathcal{X}\}$$
$$= \alpha + a_3 \hat{f}(\hat{\tau}_\alpha) + n^{-3}\{C_2(n)\hat{\gamma}_2(n) + C_3(n)\hat{\gamma}_3(n)\} + O_p(n^{-(19/5)+\gamma} + a_3^2) \quad (4.165)$$

for any $\gamma > 0$, uniformly in $|a_3| \le n^{-\epsilon_1}$. Here the smoothed bootstrap predictand $X^\dagger$ denotes, conditional on $\mathcal{X}$, a random variable which is independent of the smoothed bootstrap sample $\mathcal{X}^\dagger$ but drawn from the same population,

$$\hat{\gamma}_2(n) = E\left[\frac{\hat{f}^{(1)}(X^\dagger_{m-1:n})^2}{\hat{f}(X^\dagger_{m-1:n})^4}\{1 - \hat{F}(X^\dagger_{m-1:n})\}^3 \mathbf{I}_\mathcal{E} \,\middle|\, \mathcal{X}\right],$$

and

$$\hat{\gamma}_3(n) = E\left[\frac{\hat{f}^{(2)}(X^\dagger_{m-1:n})}{\hat{f}(X^\dagger_{m-1:n})^3}\{1 - \hat{F}(X^\dagger_{m-1:n})\}^3 \mathbf{I}_\mathcal{E} \,\middle|\, \mathcal{X}\right].$$

An estimator $\hat{a}_3$ of the additive tuner a_3 (see Subsection 4.2.2) obtained using the smoothed bootstrap is given by

$$\hat{a}_3 = \hat{G}^{-1}(\alpha) = \inf\{a_3 : \hat{G}(a_3) \ge \alpha\},$$

where $\hat{G}(a_3) = P(X^\dagger \le \hat{\xi}^\dagger_{\mathrm{Q3},\alpha} + a_3 \,|\, \mathcal{X})$. Using the expansion given by (4.165) it follows that

$$\hat{a}_3 = -n^{-3}\{c_2(n)\hat{\gamma}_2(n) + c_3(n)\hat{\gamma}_3(n)\}\hat{f}(\hat{\tau}_\alpha)^{-1} + O_p(n^{-(19/5)+\gamma}) \quad (4.166)$$

for any $\gamma > 0$.

Upon replacing a_3 by $\hat{a}_3$ on the left hand side of (4.137), conjoining the sample subset $\mathcal{E}$ with the population subset $\mathcal{A}_n$, and since the expectation of the $O_p(n^{-(19/5)+\gamma})$ term at (4.166) is $O(n^{-(19/5)+\gamma})$, it can be deduced that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}) = E\{F(\hat{\xi}_{\mathrm{Q3},\alpha} + \hat{a}_3)(\mathbf{I}_{\mathcal{A}_n \cap \mathcal{E}} + \mathbf{I}_{\mathcal{A}_n^c \cup \mathcal{E}^c})\} = \alpha + T + O(n^{-(19/5)+\gamma}) \quad (4.167)$$

for any $\gamma > 0$, where

$$T = E(\hat{a}_3)f(\xi_\alpha) + n^{-3}\{c_2(n)\gamma_2(n) + c_3(n)\gamma_3(n)\}.$$

Explicit calculation of the order of T will show additive calibration of $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, using the smoothed bootstrap, has been successful in reducing the order of coverage error. The order of T is obtained

by investigating

$$n^{-3}\{c_2(n)\zeta_2(n) + c_3(n)\zeta_3(n)\}$$

where

$$\zeta_2(n) = \gamma_2(n) - f(\xi_\alpha)E\left[\frac{\hat{f}^{(1)}(X^\dagger_{m-1:n})^2}{\hat{f}(\hat{\tau}_\alpha)\hat{f}(X^\dagger_{m-1:n})^4}\{1-\hat{F}(X^\dagger_{m-1:n})\}^3\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\right],$$

$$\zeta_3(n) = \gamma_3(n) - f(\xi_\alpha)E\left[\frac{\hat{f}^{(2)}(X^\dagger_{m-1:n})}{\hat{f}(\hat{\tau}_\alpha)\hat{f}(X^\dagger_{m-1:n})^3}\{1-\hat{F}(X^\dagger_{m-1:n})\}^3\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\right],$$

and $\gamma_2(n)$, $\gamma_3(n)$ are given by (4.133), (4.134), respectively, with $\mathcal{A}_{1,n}$ replaced by $\mathcal{A}_n \cap \mathcal{E}$.

By Taylor expansion of $\hat{f}(\hat{\tau}_\alpha)^{-1} = \{f(\hat{\tau}_\alpha) + \hat{f}(\hat{\tau}_\alpha) - f(\hat{\tau}_\alpha)\}^{-1}$, observe that

$$\hat{f}(\hat{\tau}_\alpha)^{-1} = \sum_{i=0}^{r-1}(-1)^i f(\hat{\tau}_\alpha)^{-(i+1)}\{\hat{f}(\hat{\tau}_\alpha) - f(\hat{\tau}_\alpha)\}^i + R_n \tag{4.168}$$

where, by employing bounds (4.142) and (4.152), the remainder term R_n is $O[\{h^2 + (nh)^{-1}\}^{r/2}]$ on $\mathcal{E}_2$ and can be made as small as required by selecting r sufficiently large.

Employing the expansion given by (4.168) in conjunction with similar expansions for $\hat{f}(X^\dagger_{m-1:n})^{-4}$ and $\hat{f}(X^\dagger_{m-1:n})^{-3}$ it follows that the order of T is obtained by investigating

$$\zeta_2'(n) = \gamma_2(n) - f(\xi_\alpha)E\Bigg(\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\Big[\hat{f}^{(1)}(X^\dagger_{m-1:n})^2\{1-\hat{F}(X^\dagger_{m-1:n})\}^3 \sum_{i=0}^{r-1}(-1)^i f(\hat{\tau}_\alpha)^{-(i+1)}\{\hat{f}(\hat{\tau}_\alpha) - f(\hat{\tau}_\alpha)\}^i \sum_{j=0}^{r-1}(-1)^j C_{1,j} f(X^\dagger_{m-1:n})^{-(i+4)}\{\hat{f}(X^\dagger_{m-1:n}) - f(X^\dagger_{m-1:n})\}^{j-1}\Big]\Bigg) \tag{4.169}$$

and

$$\zeta_3'(n) = \gamma_3(n) - f(\xi_\alpha)E\Bigg(\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\Big[\hat{f}^{(2)}(X^\dagger_{m-1:n})\{1-\hat{F}(X^\dagger_{m-1:n})\}^3 \sum_{i=0}^{r-1}(-1)^i f(\hat{\tau}_\alpha)^{-(i+1)}\{\hat{f}(\hat{\tau}_\alpha) - f(\hat{\tau}_\alpha)\}^i \sum_{j=0}^{r-1}(-1)^j C_{2,j} f(X^\dagger_{m-1:n})^{-(j+3)}\{\hat{f}(X^\dagger_{m-1:n}) - f(X^\dagger_{m-1:n})\}^j\Big]\Bigg), \tag{4.170}$$

where $C_{1,j}$ and $C_{2,j}$ for $j = 0, 1, \ldots, r-1$ are real constants that do not alter the derived bounds.

As shown by Falk and Reiss (1989), Taylor expansion can be used to write

$$\begin{aligned}\hat{F}(\xi_\alpha) - F(\xi_\alpha) &= \hat{F}(\xi_\alpha) - \hat{F}(\hat{\tau}_\alpha)\\ &= \hat{f}\{\xi_\alpha + \nu(\hat{\tau}_\alpha - \xi_\alpha)\}(\xi_\alpha - \hat{\tau}_\alpha),\end{aligned}$$

or equivalently,

$$\hat{\tau}_\alpha - \xi_\alpha = \frac{-1}{\hat{f}\{\xi_\alpha + \nu(\hat{\tau}_\alpha - \xi_\alpha)\}}\{\hat{F}(\xi_\alpha) - F(\xi_\alpha)\}, \tag{4.171}$$

where $0 < \nu < 1$. From integration by parts and Taylor expansion it can be shown that

$$\begin{aligned}\hat{F}(\xi_\alpha) - F(\xi_\alpha) &= \hat{F}(\xi_\alpha) - E\{\hat{F}(\xi_\alpha)\} + E\{\hat{F}(\xi_\alpha)\} - F(\xi_\alpha) \\ &= \int K(x)\{F_n(\xi_\alpha - hx) - F(\xi_\alpha - hx)\}\, dx + O(h^2).\end{aligned}$$

The first term on the right hand side may be written equivalently as

$$W = \int K(x)Y(\xi_\alpha - hx)\, dx,$$

where $Y(y) = \sum_{i=1}^n \frac{1}{n}\{\mathbf{I}_{(-\infty,y]}(X_i) - F(y)\}$. Hence it can be deduced (Fuller, 1976, Theorem 5.4.1) that

$$E(W^l) = \int \cdots \int E\Big\{\prod_{j=1}^{l} Y(\xi_\alpha - hx_i)\Big\}\prod_{i=1}^{l} K(x_i)\, dx_1 \ldots dx_l = O(n^{-k}) \tag{4.172}$$

whenever $l = 2k$ or $l = 2k - 1$ with $k \in \mathbb{N}$.

Using (4.171) and (4.172) it can be deduced that

$$E\{|\hat{\tau}_\alpha - \xi_\alpha|^l \mathbf{I}_{\mathcal{A}_n \cap \mathcal{E}}\} = O(n^{-k} + h^{2l}) \tag{4.173}$$

whenever $l = 2k$ or $l = 2k - 1$ with $k \in \mathbb{N}$. Lyapunov's inequality is required when $l = 2k - 1$.

When $l = 2$ it can be concluded from (4.172) (Bishop, Fienberg, and Holland, 1975, Theorem 14.4-1; Brockwell and Davis, 1987, Proposition 6.2.3) that $W = O_p(n^{-1/2})$. This result was used by Falk and Reiss (1989) to show that

$$\hat{\tau}_\alpha + \xi_\alpha = O_p(n^{-1/2} + h^2).$$

In general, using Markov's inequality, it can be shown that if $r > 0$, $E(X) = 0$, and $E(|X|^r) = a_n < \infty$ then $X = O_p(a_n^{-1/r})$.

Furthermore, using the Bahadur representation of order statistics (Serfling, 1980, Section 2.5.2; Bahadur, 1966) in conjunction with an analogous technique used to derive (4.172), it can be deduced that

$$E\{\mathbf{I}_{\mathcal{A}_n \cap \mathcal{E}}(X_{m-1:n} - \xi_\alpha)^l\} = O(n^{-k}) \tag{4.174}$$

and

$$E\{\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}(X^{\dagger}_{m-1:n}-\hat{\tau}_\alpha)^l\}=O(n^{-k}) \tag{4.175}$$

whenever $l=2k$ or $l=2k-1$ with $k\in\mathbb{N}$.

To derive (4.174) note that since $m-1/n=\alpha+O(n^{-1})$ it follows from Bahadur representation that

$$X_{m-1:n}-\xi_\alpha=\frac{1}{n}\sum_{i=1}^{n}Y_i+R_n \tag{4.176}$$

where $Y_i=\{\alpha-\mathbf{I}_{(-\infty,\xi_\alpha]}(X_i)\}/f(\xi_\alpha)$ and, with probability one, $R_n=O(n^{-1/2})$. Therefore $X_{m-1:n}-\xi_\alpha$ is written as a mean term plus an error term. The expectation of the mean term can be bounded using the method of Fuller (1976, Theorem 5.4.1). Therefore, using the Bahadur expansion given by (4.176) in conjunction with the c_r-inequality (Prakasa Rao, 1987, Proposition 1.14.12), it follows that (4.174) is satisfied whenever $l=2k$. Lyapunov's inequality is required to show (4.174) when $l=2k-1$. The derivation of (4.175) is similar except the tower property of conditional expectation is employed.

Observe that, with $h\sim cn^{-1/5}$,

$$\hat{f}(\hat{\tau}_\alpha)=\hat{f}(\xi_\alpha)+\sum_{j=1}^{2}\frac{1}{j}(\hat{\tau}_\alpha-\xi_\alpha)^j\hat{f}^{(j)}(\xi_\alpha)+O_p(n^{-1+\gamma})$$

for any $\gamma>0$. Therefore the order of T can be evaluated by employing the above expansion in conjunction with further Taylor expansions for $\hat{F}(X^{\dagger}_{m-1:n})$, $\hat{f}^{(1)}(X^{\dagger}_{m-1:n})$, $\hat{f}^{(2)}(X^{\dagger}_{m-1:n})$, and $f(X^{\dagger}_{m-1:n})$, all taken about ξ_α, into (4.169) and (4.170), and further substitute Taylor expansions for $F(X_{m-1:n})$, $f(X_{m-1:n})$, $f^{(1)}(X_{m-1:n})$, and $f^{(2)}(X_{m-1:n})$ into (4.169) and (4.170), all taken about ξ_α, where $\gamma_1(n)$, $\gamma_2(n)$ are given by (4.133), (4.134), respectively, with $\mathcal{A}_{1,n}$ replaced by $\mathcal{A}_n\cap\mathcal{E}$. The former Taylor expansions are terminated at a term of even order.

After this expansion the leading term of $\zeta_2'(n)$ is given by

$$f^{(1)}(\xi_\alpha)^2\{1-F(\xi_\alpha)\}^3f(\xi_\alpha)^{-3}-f(\xi_\alpha)^{-3}E\Big[\hat{f}^{(1)}(\xi_\alpha)^2\{1-\hat{F}(\xi_\alpha)\}^3\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\Big]. \tag{4.177}$$

Proceed by writing

$$E\Big[\hat{f}^{(1)}(\xi_\alpha)^2\{1-\hat{F}(\xi_\alpha)\}^3\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\Big]=E\Big(\big[\hat{f}^{(1)}(\xi_\alpha)-E\{\hat{f}^{(1)}(\xi_\alpha)\}+E\{\hat{f}^{(1)}(\xi_\alpha)\}\big]^2$$
$$\big[1-\hat{F}(\xi_\alpha)-1+E\{\hat{F}(\xi_\alpha)\}+1-E\{\hat{F}(\xi_\alpha)\}\big]^3\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\Big).$$

Expand the above term, note the moments about the mean are given by (4.143) and (4.144) with the bias of these terms given by (4.153) and (4.158). By employing Hölder's and Lyapunov's inequalities it follows that (4.177) can be written as

$$O\{h^2 + (nh^3)^{-1}\}. \tag{4.178}$$

Using (4.174) and since the population distribution function F has five bounded derivatives with the first derivative bounded away from zero at ξ_α, the remainder term obtained from the Taylor expansion of $\gamma_2(n)$ is given by

$$O(n^{-1}). \tag{4.179}$$

Additionally the remainder term obtained from the smoothed bootstrap components of (4.169) can be shown to be given by

$$O\{(nh^3)^{-1}\}. \tag{4.180}$$

From (4.178), (4.179), and (4.180) it follows that $\zeta_1'(n)$ can be replaced by $O\{h^2 + (nh^3)^{-1}\}$. Analogously $\zeta_2'(n)$ can be replaced by $O\{h^2 + (nh^3)^{-1}\}$. Therefore it follows that

$$T = O\big[n^{-3}\{h^2 + (nh^3)^{-1}\}\big] \tag{4.181}$$

where the h^2 term comes from bias and the $(nh^3)^{-1}$ term comes from the components involving $\hat{f}^{(1)}(\xi_\alpha)^2$ and $\hat{f}(\xi_\alpha)\hat{f}^{(2)}(\xi_\alpha)$. Since $h \sim cn^{-1/5}$, replace (4.181) by $T = O(n^{-3-2/5})$. Hence, (4.114) follows.

Consider now the coverage error properties of $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha} + \hat{a}_2]$, where

$$\hat{a}_2 = \hat{G}^{-1}(\alpha) = \inf\{a_2 : \hat{G}(a_2) \geq \alpha\},$$

and

$$\hat{G}(a_2) = P(X^\dagger \leq \hat{\xi}^\dagger_{\mathrm{Q2},\alpha} + a_2 \,|\, \mathcal{X}).$$

From technical arguments analogous to those used when deriving (4.167), in conjunction with a longer Taylor expansion, it can be shown that

$$P(X \in \widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}) = \alpha + T + o(n^{-8/3}).$$

In this case

$$T = n^{-2}c_1(n)\left[\gamma_1(n) - E\{\hat{\gamma}_1(n)/\hat{f}(\hat{\tau}_\alpha)\}f(\xi_\alpha)\mathbf{I}_{\mathcal{A}_n\cap\mathcal{E}}\right]$$

where $c_1(n)$ is bounded in n,

$$\gamma_1(n) = E\left[\frac{f^{(1)}(X_{m:n})}{f(X_{m:n})^2}\{1 - F(X_{m:n})\}^2\mathbf{I}_{\mathcal{A}_n}\right],$$

and

$$\hat{\gamma}_1(n) = E\left[\frac{\hat{f}^{(1)}(X^\dagger_{m:n})}{\hat{f}(X^\dagger_{m:n})}\{1 - \hat{F}(X_{m:n})\}^2\mathbf{I}_{\mathcal{E}} \,\middle|\, \mathcal{X}\right].$$

From technical arguments analogous to those used when determining (4.181) it can be shown that

$$T = O\{n^{-2}(h^2 + t)\},$$

where the h^2 term comes from bias and t denotes an upper bound to the order of the variance of $\hat{f}(\xi_\alpha)$ and the covariance of $\hat{f}(\xi_\alpha)$ and $\hat{f}^{(1)}(\xi_\alpha)$. The former is of order $(nh)^{-1}$; the latter is of order

$$(nh)^{-1}(nh^2)^{-1}\left\{n\left|\int K^{(1)}\left(\frac{x-y}{h}\right)K\left(\frac{x-y}{h}\right)dy\right| + nh\int\left|K^{(1)}\left(\frac{x-y}{h}\right)K\left(\frac{x-y}{h}\right)\right|dy\right\} = O\{(nh)^{-1}\},$$

since $\int K^{(1)}(t)K(t)\,dt = 0$ by integration by parts. When $h \sim cn^{-1/3}$ it follows that

$$T = O(n^{-2-2/3}).$$

Hence, (4.113) follows. This concludes the proof. ∎

Chapter 5

Numerical Evaluation and Properties of Prediction Intervals

The purpose of this chapter is, firstly, to elucidate various numerical obscurities of the nominal one-sided α-level prediction intervals given in Chapter 4; secondly, to investigate the numerical properties of prediction intervals via a simulation study. A numerical investigation delivers tentative and anecdotal answers to questions that may plague a statistician's mind when employing a nominal prediction interval. These include:

- Given a nominal prediction interval and $\epsilon > 0$, how large does sample size have to be before the absolute value of coverage error is bounded by ϵ?
- For a fixed sample size, what is the coverage error of the nominal predication interval?
- For a fixed sample size, how does coverage error of the nominal prediction interval depend on the nominal coverage?

Precise answers to the former questions are not possible from the asymptotic coverage error specifications in Chapter 4. By considering numerous parametric models and computing coverage error the indicative behaviour for a nominal prediction interval can be assessed. Furthermore, the computed coverage errors may be used to evaluate relative performance of the competing nominal prediction intervals.

An outline of the present chapter is as follows. Section 5.1 elaborates the numerical approximation used for coverage error, Section 5.2 elucidates the numerical approximation used for the

additive calibration of a nominal prediction interval, and Section 5.3 documents the indicative but anecdotal properties for all types of prediction intervals given in Chapter 4, using a simulation study.

5.1 Coverage Error Approximation

Let f denote a probability density function and $F(x) = \int_{-\infty}^{x} f(y)\,dy$ the corresponding probability distribution function. Set $\alpha \in (0,1)$ and let $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ denote n independent and identically distributed random variables drawn from a population with F as probability distribution function. Denote a generic nominal one-sided α-level prediction interval by

$$\widehat{\mathcal{I}}_\alpha = (-\infty, \hat{\xi}_\alpha],$$

where the upper end-point $\hat{\xi}_\alpha$ is an estimator of the α-th population quantile $\xi_\alpha = \inf\{\xi : F(\xi) \geq \alpha\}$ and depends on α, the sample size n, and the sample $\mathcal{X}$ only. The coverage error of $\widehat{\mathcal{I}}_\alpha$ is given by

$$\kappa(\alpha, n) = P(X \in \widehat{\mathcal{I}}_\alpha) - \alpha,$$

where X is a random variable which is independent of the sample $\mathcal{X}$ but drawn from the same population.

Using the joint probability density of order statistics (Reiss, 1989, Theorem 1.5.6), it follows that $\kappa(\alpha, n)$ can be approximated via direct numerical integration.

Alternatively, an estimator of $\kappa(\alpha, n)$ can be constructed as follows. Draw B_1 independent versions of the sample $\mathcal{X}$ which will be denoted by

$$\begin{aligned}
\mathcal{X}_1 &= \{X_{1,1}, X_{1,2}, \ldots, X_{1,n}\},\\
\mathcal{X}_2 &= \{X_{2,1}, X_{2,2}, \ldots, X_{2,n}\},\\
&\vdots\\
\mathcal{X}_{B_1} &= \{X_{B_1,1}, X_{B_1,2}, \ldots, X_{B_1,n}\}.
\end{aligned}$$

For each $k = 1, 2, \ldots, B_1$ calculate $\hat{\xi}_{k,\alpha}$ which denotes the version of $\hat{\xi}_\alpha$ computed using the k-th sample $\mathcal{X}_k$ instead of the sample $\mathcal{X}$. Then an estimator of $\kappa(\alpha, n)$ is given by

$$\hat{\kappa}_{B_1}(\alpha, n) = B_1^{-1} \sum_{k=1}^{B_1} \{F(\hat{\xi}_{k,\alpha}) - \alpha\}. \tag{5.1}$$

Using the tower property of conditional expectation (see (2.12)),

$$P(X \in \widehat{\mathcal{I}}_\alpha) = E\{\mathbf{I}_{\widehat{\mathcal{I}}_\alpha}(X)\} = E\Big[E\{\mathbf{I}_{\widehat{\mathcal{I}}_\alpha}(X) \mid \mathcal{X}\}\Big] = E\{F(\hat{\xi}_\alpha)\}.$$

The former expression in conjunction with the strong law of large numbers (Chung, 1974; Shiryayev, 1984) can be used to establish that, conditional on $\mathcal{X}$, $\hat{\kappa}_{B_1}(\alpha, n) \xrightarrow{\text{w.p.1}} \kappa(\alpha, n)$ as $B_1 \to \infty$, where $\xrightarrow{\text{w.p.1}}$ denotes convergence with probability one.

Techniques for generating samples with probability distribution function F can, for example, be found in Ripley (1987).

When $\widehat{\mathcal{I}}_\alpha$ denotes $\widehat{\mathcal{I}}_{\text{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\text{Q2},\alpha}$ or $\widehat{\mathcal{I}}_{\text{Q3},\alpha}$ (see Section 4.1) then $\hat{\xi}_\alpha$ denotes $\hat{\xi}_{\text{Q1},\alpha}$, $\hat{\xi}_{\text{Q2},\alpha}$ or $\hat{\xi}_{\text{Q3},\alpha}$, respectively, and the coverage error estimator given by (5.1) can be utilised directly. However, when $\widehat{\mathcal{I}}_\alpha$ denotes $\widehat{\mathcal{I}}_{\text{AC,Q2},\alpha}$ or $\widehat{\mathcal{I}}_{\text{AC,Q3},\alpha}$ (see Section 4.1) then $\hat{\xi}_\alpha$ denotes $\hat{\xi}_{\text{Q2},\alpha} + \hat{a}_2$ or $\hat{\xi}_{\text{Q3},\alpha} + \hat{a}_3$, respectively, and the coverage error estimation given by (5.1) can be utilised directly only after the additive tuner estimators $\hat{a}_2$ and $\hat{a}_3$ have been computed for each of the k-th samples $\mathcal{X}_k$ for $k = 1, 2, \ldots, B_1$. The numerical computation of additive tuner estimators is considered in the next section.

5.2 Approximation of Additive Tuner Estimators

Let $\hat{f}$ denote the kernel density estimator (see Subsection 4.2.2) of the population probability density function f which is based on the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ which denotes n independent and identically distributed random variables drawn from a population with probability distribution function F. Let the smoothed bootstrap sample $\mathcal{X}^\dagger = (X_1^\dagger, X_2^\dagger, \ldots, X_n^\dagger)^\top$ denote, conditional on $\mathcal{X}$, n independent and identically distributed random variables drawn from a population with $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(y)\, dy$ as probability distribution function. For fixed $\alpha \in (0,1)$ the additive-calibrated versions of $\widehat{\mathcal{I}}_{\text{Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\text{Q3},\alpha}$ derived using the smoothed bootstrap are given by $\widehat{\mathcal{I}}_{\text{AC,Q2},\alpha} = (-\infty, \hat{\xi}_{\text{Q2},\alpha} + \hat{a}_2]$ and $\widehat{\mathcal{I}}_{\text{AC,Q3},\alpha} = (-\infty, \hat{\xi}_{\text{Q3},\alpha} + \hat{a}_3]$, respectively. Here $\hat{\xi}_{\text{Q2},\alpha}$ and $\hat{\xi}_{\text{Q3},\alpha}$ are given by (4.10) and (4.17), respectively, and the additive tuner estimators $\hat{a}_2$ and $\hat{a}_3$ are the solutions of $\hat{p}_2(a_2) = \alpha$ and $\hat{p}_3(a_3) = \alpha$, respectively, where

$$\hat{p}_2(a_2) = P(X^\dagger \in \mathcal{I}^\dagger_{\text{AC,Q2},\alpha} \mid \mathcal{X}) \quad \text{and} \quad \hat{p}_3(a_3) = P(X^\dagger \in \mathcal{I}^\dagger_{\text{AC,Q3},\alpha} \mid \mathcal{X}),$$

$\mathcal{I}^\dagger_{\text{AC,Q2},\alpha} = (-\infty, \hat{\xi}^\dagger_{\text{Q2},\alpha} + a_2]$ and $\mathcal{I}^\dagger_{\text{AC,Q3},\alpha} = (-\infty, \hat{\xi}^\dagger_{\text{Q3},\alpha} + a_3]$ with $\hat{\xi}^\dagger_{\text{Q2},\alpha}$ and $\hat{\xi}^\dagger_{\text{Q3},\alpha}$ denoting the versions of $\hat{\xi}_{\text{Q2},\alpha}$ and $\hat{\xi}_{\text{Q3},\alpha}$, respectively, calculated using the smoothed bootstrap sample $\mathcal{X}^\dagger$

instead of the sample $\mathcal{X}$. The smoothed bootstrap predictand $X^\dagger$ denotes, conditional on $\mathcal{X}$, a random variable which is independent of $\mathcal{X}^\dagger$ but drawn from the same population.

Draw B_2 independent versions of $\mathcal{X}^\dagger$ which will be denoted by

$$\begin{aligned}\mathcal{X}_1^\dagger &= \{X_{1,1}^\dagger, X_{1,2}^\dagger, \ldots, X_{1,n}^\dagger\},\\ \mathcal{X}_2^\dagger &= \{X_{2,1}^\dagger, X_{2,2}^\dagger, \ldots, X_{2,n}^\dagger\},\\ &\vdots\\ \mathcal{X}_{B_2}^\dagger &= \{X_{B_2,1}^\dagger, X_{B_2,2}^\dagger, \ldots, X_{B_2,n}^\dagger\}.\end{aligned}$$

For each $k = 1, 2, \ldots, B_2$, calculate $\hat{\xi}_{k,\mathrm{Q2},\alpha}^\dagger$ and $\hat{\xi}_{k,\mathrm{Q3},\alpha}^\dagger$ which are versions of $\hat{\xi}_{\mathrm{Q2},\alpha}^\dagger$ and $\hat{\xi}_{\mathrm{Q3},\alpha}^\dagger$, respectively, computed using the k-th smoothed bootstrap sample $\mathcal{X}_k^\dagger$ instead of the smoothed bootstrap sample $\mathcal{X}^\dagger$. Estimators for $\hat{p}_2(a_2)$ and $\hat{p}_3(a_3)$ are then given by

$$\hat{p}_{B_2,2}(a_2) = B_2^{-1} \sum_{k=1}^{B_2} \hat{F}(\hat{\xi}_{k,\mathrm{Q2},\alpha}^\dagger + a_2)$$

and

$$\hat{p}_{B_2,3}(a_3) = B_2^{-1} \sum_{k=1}^{B_2} \hat{F}(\hat{\xi}_{k,\mathrm{Q3},\alpha}^\dagger + a_3),$$

respectively. Approximations for the additive tuner estimators $\hat{a}_2$ and $\hat{a}_3$ derived via the smoothed bootstrap are then given by

$$\hat{a}_{B_2,2} = \hat{p}_{B_2,2}^{-1}(\alpha) \qquad \text{and} \qquad \hat{a}_{B_2,3} = \hat{p}_{B_2,3}^{-1}(\alpha), \tag{5.2}$$

respectively, where $\hat{p}_{B_2,2}^{-1}(\alpha) = \inf\{a : \hat{p}_{B_2,2}(a) \geq \alpha\}$ and $\hat{p}_{B_2,3}^{-1}(\alpha) = \inf\{a : \hat{p}_{B_2,3}(a) \geq \alpha\}$. In practice, a grid technique in conjunction with a root finding routine can be used to numerically identify $\hat{a}_{B_2,2}$ and $\hat{a}_{B_2,3}$.

The strong law of large numbers can be used to establish that, conditional on $\mathcal{X}$, $\hat{p}_{B_2,2}(a_2) \xrightarrow{\text{w.p.1}} \hat{p}_2(a_2)$ and $\hat{p}_{B_2,3}(a_3) \xrightarrow{\text{w.p.1}} \hat{p}_3(a_3)$ as $B_2 \to \infty$. Using the former result in conjunction with the techniques illustrated in Serfling (1980, Section 2.3.1) it follows that for each $\epsilon > 0$,

$$P\left(\sup_{B \geq B_2} |\hat{a}_{B,2} - \hat{a}_2| > \epsilon \,\middle|\, \mathcal{X} \right) \to 0$$

and

$$P\left(\sup_{B \geq B_2} |\hat{a}_{B,3} - \hat{a}_3| > \epsilon \,\middle|\, \mathcal{X} \right) \to 0,$$

as $B_2 \to \infty$. Hence, conditional on $\mathcal{X}$, $\hat{a}_{B_2,2} \xrightarrow{\text{w.p.1}} \hat{a}_2$ and $\hat{a}_{B_2,3} \xrightarrow{\text{w.p.1}} \hat{a}_3$ as $B_2 \to \infty$.

Combining (5.1) and (5.2) it follows that the coverage error approximations for the additive-calibrated prediction intervals $\widehat{\mathcal{I}}_{\text{AC,Q2},\alpha} = (-\infty, \hat{\xi}_{\text{Q2},\alpha} + \hat{a}_2]$ and $\widehat{\mathcal{I}}_{\text{AC,Q3},\alpha} = (-\infty, \hat{\xi}_{\text{Q3},\alpha} + \hat{a}_3]$ are given by

$$\hat{\kappa}_{B_1,B_2,2}(\alpha, n) = B_1^{-1} \sum_{k=1}^{n} \left\{ F(\hat{\xi}_{k,\text{Q2},\alpha} + \hat{a}_{k,B_2,2}) - \alpha \right\} \tag{5.3}$$

and

$$\hat{\kappa}_{B_1,B_2,3}(\alpha, n) = B_1^{-1} \sum_{k=1}^{n} \left\{ F(\hat{\xi}_{k,\text{Q3},\alpha} + \hat{a}_{k,B_2,3}) - \alpha \right\} \tag{5.4}$$

respectively, where $\hat{a}_{k,B_2,2}$ and $\hat{a}_{k,B_2,3}$ denote versions of $\hat{a}_{B_2,2}$ and $\hat{a}_{B_2,3}$ computed using the sample $\mathcal{X}_k$ for $k = 1, 2, \ldots, B_1$ instead of the sample $\mathcal{X}$.

Since $F(\hat{\xi}_{k,\text{Q2},\alpha} + \hat{a}_{k,B_2,2})$ and $F(\hat{\xi}_{k,\text{Q3},\alpha} + \hat{a}_{k,B_2,3})$ are bounded, hence uniformly integrable, it follows from the mean convergence criteria (Chow and Teicher, 1997) that, conditional on $\mathcal{X}$, $\hat{\kappa}_{B_1,B_2,2}(\alpha, n) \xrightarrow{\text{w.p.1}} \hat{\kappa}_{\text{AC},2}(\alpha, n)$ and $\kappa_{B_1,B_2,3}(\alpha, n) \xrightarrow{\text{w.p.1}} \kappa_{\text{AC},3}(\alpha, n)$ as B_1, $B_2 \to \infty$, where

$$\kappa_{\text{AC},2}(\alpha, n) = P(X \in \widehat{\mathcal{I}}_{\text{AC,Q2},\alpha}) - \alpha$$

and

$$\kappa_{\text{AC},3}(\alpha, n) = P(X \in \widehat{\mathcal{I}}_{\text{AC,Q3},\alpha}) - \alpha.$$

Lee and Young (1999) investigate the coverage error properties of an α-level confidence interval derived from Monte-Carlo approximations; it is expected that an analogous results would hold for a prediction interval constructed using the former approximations.

Smoothed Bootstrap Sample Generation. Conditional on $\mathcal{X}$, the realisation of a smoothed bootstrap sample $\mathcal{X}^\dagger$ can be generated without any reference to the kernel density estimator $\hat{f}$. Silverman (1986, Section 6.4.1) presents an algorithm for generation of a sample $\mathcal{X}^\dagger$ which, in the notation of Golub and Van Loan (1996), proceeds as follows.

Algorithm 5.5 *Denote the sample vector of length n by $\mathcal{X}(1{:}n)$. Set $\mathcal{X}(i) = X_i$ and let h denote the bandwidth used to construct the kernel estimator $\hat{f}$ (see (4.40)). Conditional on $\mathcal{X}$, this algorithm generates a smoothed bootstrap sample vector of length n given by $\mathcal{X}^\dagger(1{:}n)$, from a*

population with probability density function $\hat{f}$. *The algorithm relies on two predefined functions:* **uniform_index_realisation**(i, j) *generates an index uniformly with replacement from* $i{:}j$, *where* $i < j$ *and* $i, j \in \mathbb{Z}$; **K_realisation** *generates a realisation from a population with kernel* K *as probability density function. Successive calls to* **uniform_index_realisation** *or* **K_realisation** *deliver independent realisations.*

1 $n, h, i, \mathcal{X}(1{:}n) := [X_1, X_2, \ldots, X_n]^\top, \mathcal{X}^\dagger(1{:}n)$;
2 **for** $i = 1{:}n$,
3 $\quad \mathcal{X}^\dagger(i) := \mathcal{X}(\textbf{uniform\_index\_realisation}(1, n)) + h * \textbf{K\_realisation}()$;
4 **end**

If $\mathcal{X}(\textbf{uniform\_index\_realisation}(1, n))$ is replaced by $\mathcal{X}(i)$ in line 3 of Algorithm 5.5 a Kendall bootstrap sample is generated. The Kendall bootstrap is an alternative to the smoothed bootstrap (Young, 1990).

Let the kernel K of the kernel density estimator $\hat{f}$ at (4.40), be given by the rescaled Epanechnikov kernel

$$K_{\mathrm{E}}(x) = \begin{cases} \frac{3}{4}(1 - x^2), & \text{if } |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

This kernel is well known in the data smoothing community as it is the unscaled version of this kernel that minimises the leading or asymptotic term of the mean iterated squared error,

$$\int_{-\infty}^{\infty} E\{\hat{f}(x) - f(x)\}^2 \, dx,$$

provided the bandwidth h is first selected to minimise the asymptotic mean iterated squared error. See, for example, Silverman (1986) and Simonoff (1996).

Another advantage of the rescaled Epanechnikov kernel is that it is not computationally difficult to simulate from because of an algorithm given by Devroye and Györfi (1985). In this case Algorithm 5.5 has function **K_realisation** defined as follows.

Algorithm 5.6 *This function returns a realisation of a population with rescaled Epanechnikov kernel as probability density function and relies on the predefined function* **uniform_realisation**(a, b) *which generates a random variable that is uniformly distributed on* (a, b), *where* $a < b$ *and* $a, b \in \mathbb{R}$. *Successive calls to* **uniform_realisation** *deliver independent realisations.*

```
1  function   [ek] = K_realisation();
2    a := −1, b := 1, U1, U2, U3;
3    U1 := uniform_realisation(a, b);
4    U2 := uniform_realisation(a, b);
5    U3 := uniform_realisation(a, b);
6    if {(|U3| ≥ |U2|) and (|U3| ≥ |U1|)},
7      ek := U2;
8    else
9      ek := U3;
10   end
11 end
```

5.3 Numerical Properties

In this section versions of the coverage error approximation, given by (5.1), for the prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q1},\alpha}]$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha}]$, and $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q3},\alpha}]$ instead of $\widehat{\mathcal{I}}_{\alpha} = (-\infty, \hat{\xi}_{\alpha}]$, together with coverage error approximations, given by (5.3) and (5.4), for the prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q2},\alpha} + \hat{a}_2]$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha} = (-\infty, \hat{\xi}_{\mathrm{Q3},\alpha} + \hat{a}_3]$, respectively, will be used to investigate the numerical properties for these prediction intervals.

Given the sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)^\top$ two types of bandwidths, $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$, and, $\hat{h}_{\mathrm{S,Q2},\alpha} = \hat{\sigma}\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{S,Q3},\alpha} = \hat{\sigma}\hat{h}_{\mathrm{N,Q2},\alpha}$, are propounded in Section 4.2 of Chapter 4 for constructing prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$. Here $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$; $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$ are derived from the appropriate population density $f(x\,;\hat{\theta})$ (see Section 4.2 for notational explanations) and are defined by

$$\hat{h}_{\mathrm{P,Q2},\alpha} = \arg\min_h \left| P(X' \in \widehat{\mathcal{I}}'_{\mathrm{AC,Q2},\alpha} \,|\, \mathcal{X}) - \alpha \right|, \tag{5.6}$$

$$\hat{h}_{\mathrm{P,Q3},\alpha} = \arg\min_h \left| P(X' \in \widehat{\mathcal{I}}'_{\mathrm{AC,Q3},\alpha} \,|\, \mathcal{X}) - \alpha \right|, \tag{5.7}$$

and where $\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{N,Q3},\alpha}$ denote versions of $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$, respectively, in which $f(x\,;\hat{\theta})$ is replaced by the standard normal density.

The numerical computation of $\hat{h}_{\mathrm{P,Q2},\alpha}$ and $\hat{h}_{\mathrm{P,Q3},\alpha}$ would replace (5.6) and (5.7) by

$$\hat{h}_{B_1,B_2,\mathrm{P,Q2},\alpha} = \arg\min_h \left| \hat{\kappa}_{B_1,B_2,2}(\alpha, n) \right| \tag{5.8}$$

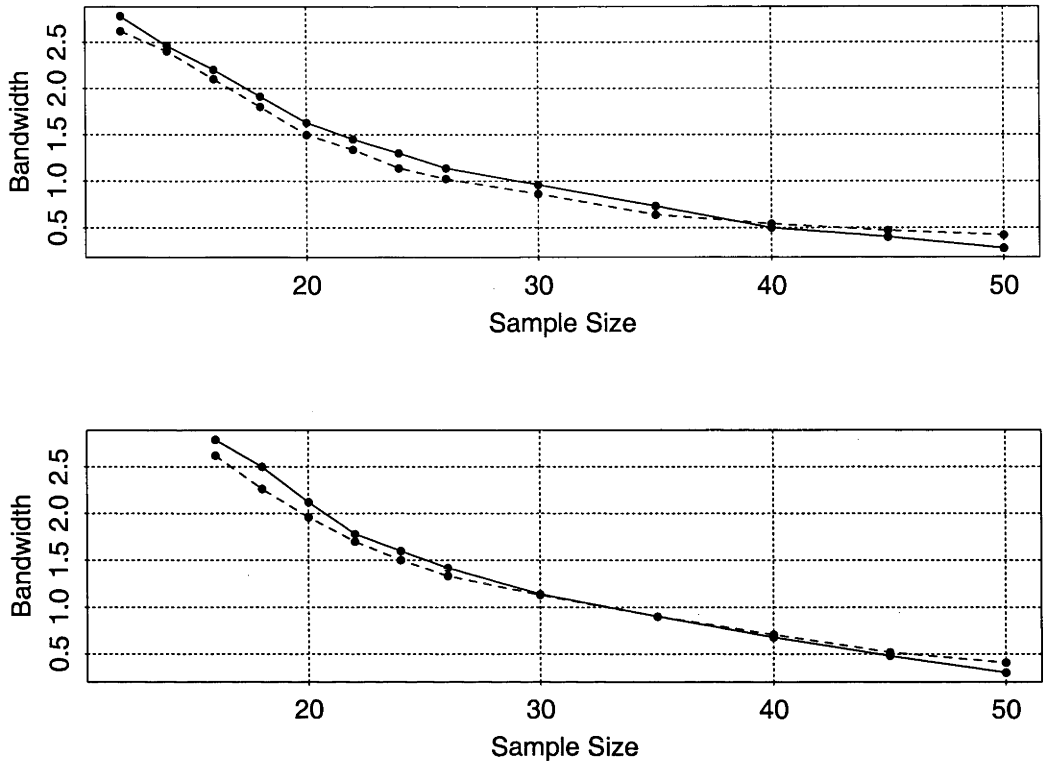

Figure 5.1: Optimal Bandwidths. Plotted are bandwidth values that minimise absolute coverage error for a standard normal population, various values of sample size, n, and nominal coverages $\alpha = 0.9$ and 0.95 for the former and latter plot, respectively. Bandwidths $\hat{h}_{\mathrm{N,Q2},\alpha}$ (dashed line) and $\hat{h}_{\mathrm{N,Q3},\alpha}$ (solid line) are appropriate for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively.

and

$$\hat{h}_{B_1,B_2,\mathrm{P,Q3},\alpha} = \arg\min_h \left|\hat{\kappa}_{B_1,B_2,3}(\alpha,n)\right|, \tag{5.9}$$

where $\hat{\kappa}_{B_1,B_2,2}(\alpha,n)$ and $\hat{\kappa}_{B_1,B_2,3}(\alpha,n)$ are versions of (5.3) and (5.4), respectively, for a population with $f(x\,;\hat{\theta})$ as probability density function.

In practice, a grid technique in conjunction with a function minimising routine would be used to identify $\hat{h}_{B_1,B_2,\mathrm{P,Q2},\alpha}$ and $\hat{h}_{B_1,B_2,\mathrm{P,Q3},\alpha}$.

Let $\hat{h}_{B_1,B_2,\mathrm{N,Q2},\alpha}$ and $\hat{h}_{B_1,B_2,\mathrm{N,Q3},\alpha}$ denote versions of (5.8) and (5.9), respectively, obtained using a standard normal density instead of $f(x\,;\hat{\theta})$. Figure 5.1 gives, for a range of different sample sizes, the approximate bandwidths of $\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{N,Q3},\alpha}$ which minimise absolute coverage error for nominal one-sided α-level prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively, for a standard normal population when $\alpha = 0.9$ or 0.95, $B_1 = 5000$, and $B_2 = 2000$. For a sample size not

considered in Figure 5.1 linear interpolation may be used. Table 1 of Appendix A can be used for this purpose.

To investigate the numerical properties of prediction intervals given in Chapter 4, four kinds of population model P are treated, corresponding to a different population probability density function $f(x;\theta_0)$:

D1 Standard normal population which has probability density function

$$f(x;\theta_0) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{x^2}{2}\right);$$

D2 Standard exponential population which has probability density function

$$f(x;\theta_0) = \begin{cases} \exp(-x) & \text{if } x \in (0,\infty) \\ 0 & \text{otherwise} \end{cases};$$

D3 Two-parameter Weibull population, with shape and scale parameters given by 2 and 1 respectively, which has probability density function

$$f(x;\theta_0) = \begin{cases} 2x\exp(-x^2) & \text{if } x \in (0,\infty) \\ 0 & \text{otherwise} \end{cases};$$

D4 Standard lognormal population which has probability density function

$$f(x;\theta_0) = \begin{cases} \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{(\log x)^2}{2}\right\} & \text{if } x \in (0,\infty) \\ 0 & \text{otherwise} \end{cases}.$$

For various sample sizes the approximate coverage errors of the nominal α-level prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, defined at (4.6), (4.11), and (4.18), respectively, and, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, defined at (4.41), are given in Figures 5.5, 5.6, and 5.7 when the nominal coverage $\alpha = 0.9$, for all four kinds of population model. A tabulated version of these results are given by Table 2 of Appendix A. Coverage error results for each population model are obtained with $B_1 = 5000$ and $B_2 = 2000$. The rescaled Epanechnikov kernel, see (5.5), is used for additive calibration since it has compact support, and an appropriate bandwidth, $\hat{h}_{\mathrm{S,Q2},\alpha}$ or $\hat{h}_{\mathrm{S,Q3},\alpha}$, is chosen by interpolation from Table 1 and scaled by $\hat{\sigma}$, as discussed in Subsection 4.2.2. Algorithm 5.5 is used in conjunction with Algorithm 5.6 to generate the B_2 smoothed bootstrap samples. Appropriate algorithms are

used to generate the B_1 population samples; see for example Ripley (1987) or the NAG fortran library[1].

The upper end-point for the prediction interval $\widehat{\mathcal{I}}_{Q3,\alpha} = (-\infty, \hat{\xi}_{Q3,\alpha}]$ is given by (4.22), where C and a, b, c are given by (4.20) and (4.21), respectively; note that $\widehat{\mathcal{I}}_{AC,Q3,\alpha}$ then denotes the additive-calibrated version of $\widehat{\mathcal{I}}_{Q3,\alpha}$ assuming the former coefficient selection. For this quadruplet (m, a, b, c), $\hat{\xi}_{Q3,\alpha}$ is nonnegative when the support of the population is nonnegative (Subsection 4.1.3).

By comparing these coverage error results it can be seen that coverage error is reduced by replacing the prediction interval $\widehat{\mathcal{I}}_{Q2,\alpha}$, derived via interpolation among two order statistics, with $\widehat{\mathcal{I}}_{Q3,\alpha}$, derived via interpolation among three order statistics. Additive calibration using the smoothed bootstrap is also an effective tool for reducing coverage error when sample size is small; see results for $\widehat{\mathcal{I}}_{AC,Q2,\alpha}$ and $\widehat{\mathcal{I}}_{AC,Q3,\alpha}$. Poor relative performance is exhibited by prediction interval $\widehat{\mathcal{I}}_{Q1,\alpha}$.

Figures 5.5, 5.6, and 5.7 illustrate coverage error properties when sample size is $n = 12$ and the population model P is given by D1, D2, D3 or D4, for various values of nominal coverages. Tabulated versions of these results are given by Tables 3, 4, 5, and 6 in Appendix A. For this investigation the bandwidth used for prediction intervals $\widehat{\mathcal{I}}_{AC,Q2,\alpha}$ and $\widehat{\mathcal{I}}_{AC,Q3,\alpha}$ is fixed for all α and given simple by $n^{-1/3}$ and $n^{-1/5}$, respectively, where n denotes sample size, with all other former specifications remaining unchanged.

For all population models P given by D1, D2, D3, D4 the prediction interval $\widehat{\mathcal{I}}_{Q1,\alpha}$ is conservative and has poor performance. In fact the coverage error fluctuates erratically in the interval $[-0.08, 0]$ as a consequence of the discontinuous nature of the integer part function which plays a prominent role in the definition of $\widehat{\mathcal{I}}_{Q1,\alpha}$ (Subsection 4.1.1). The coverage error of prediction interval $\widehat{\mathcal{I}}_{Q3,\alpha}$ is smaller than prediction interval $\widehat{\mathcal{I}}_{Q2,\alpha}$, uniformly for all values of α considered. Using the smoothed bootstrap for additive calibration is also, in general, effective in reducing the magnitude of coverage error; compare entries for $\widehat{\mathcal{I}}_{Q2,\alpha}$ and $\widehat{\mathcal{I}}_{Q3,\alpha}$ against $\widehat{\mathcal{I}}_{AC,Q2,\alpha}$ and $\widehat{\mathcal{I}}_{AC,Q3,\alpha}$, respectively, in Tables 3, 4, 5, and 6.

Using the smoothed bootstrap for additive calibration may slightly increase coverage error (see $\alpha = 0.30$ in Tables 3 and 6, and $\alpha = 0.25$ in Table 4). In the case where the population model P has nonnegative support (D2, D3, and D4) this may be an artifact of the kernel density estimator

[1] Fortran 77 library, Mark 18.

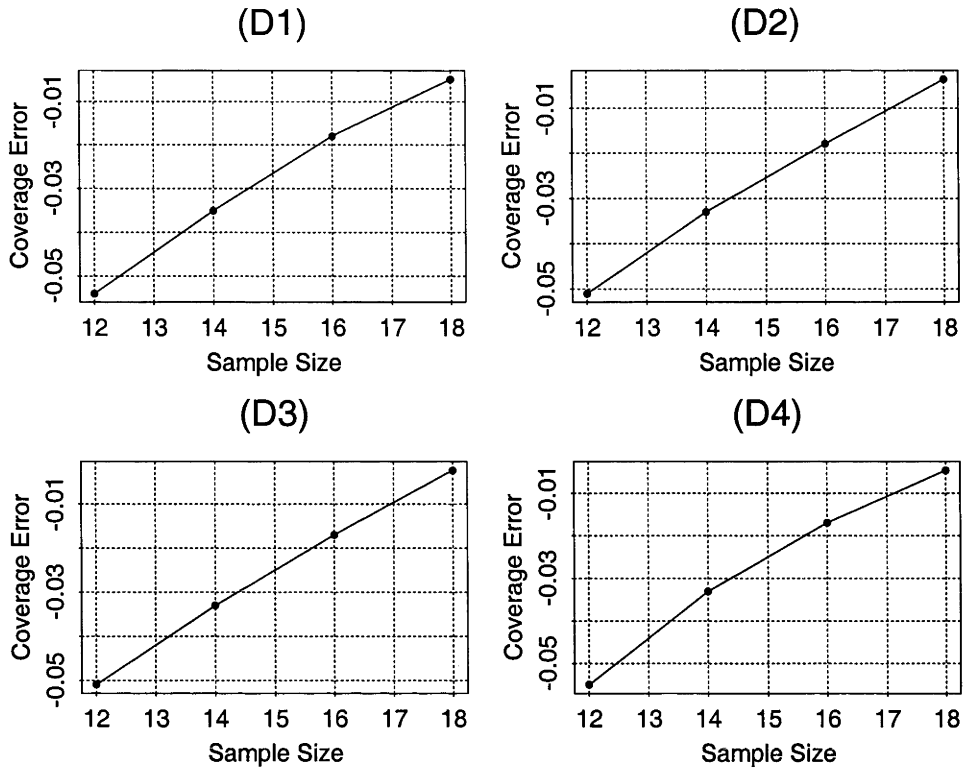

Figure 5.2: Coverage Errors. Values of coverage error for prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, when the population model P is given by D1, D2, D3, and D4, the sample size n is given by 12, 14, 16, and 18, and nominal coverage $\alpha = 0.9$.

$\hat{f}$ having spurious negative support. When the population model P has nonnegative support the use of a kernel density estimator $\hat{f}$ which has nonnegative support may deliver superior numerical results when used for additive calibration. One plausible methodology for constructing such a kernel density estimator, which is similar in style to Owen (1988) and Hall and Presnell (1999), would be to consider

$$\hat{f}_{\mathrm{p}}(x) = h^{-1} \sum_{i=1}^{n} p_i K\left(\frac{x - X_i}{h}\right),$$

where $p = (p_1, p_2, \ldots, p_n)$ is a probability distribution on the sample $\mathcal{X}$. To impose the condition that the kernel density estimator $\hat{f}_{\mathrm{p}}$ has nonnegative support, select $p = \hat{p}$ to minimise the distance $D(p)$, say, from p to $p_{\mathrm{unif}} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ subject to $\hat{f}_{\mathrm{p}}(x) = 0$ whenever $x \in (-\infty, 0)$.

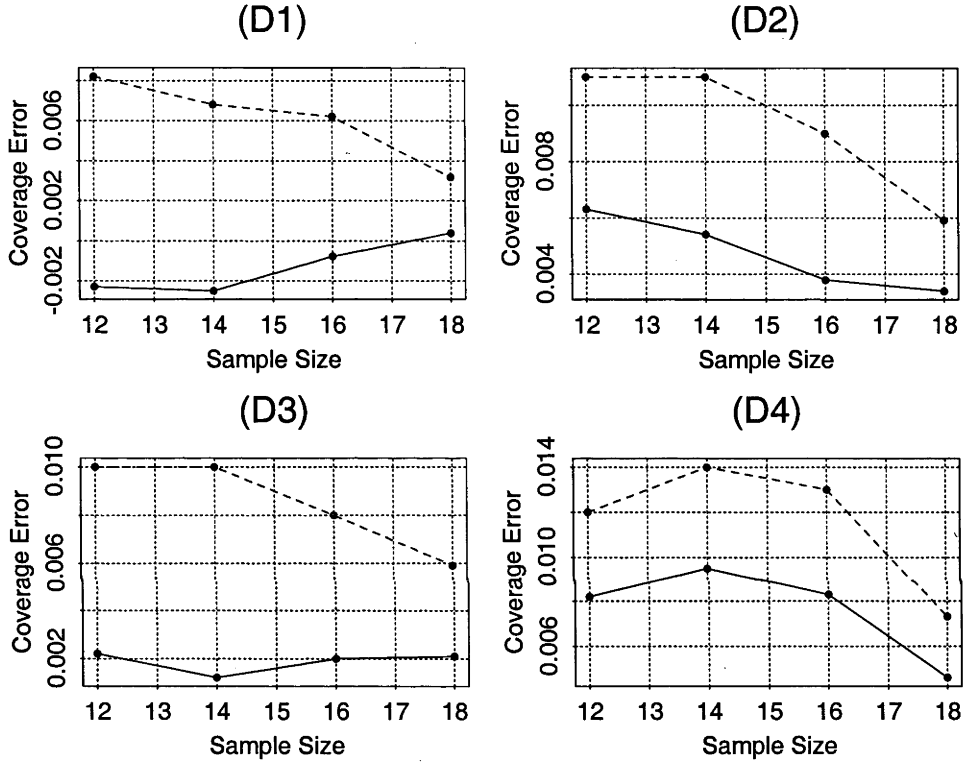

Figure 5.3: Coverage Errors. Values of coverage error for prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ (dashed line) and $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ (solid line), when the population model P is given by D1, D2, D3, and D4, the sample size n is given by 12, 14, 16, and 18, and nominal coverage $\alpha = 0.9$.

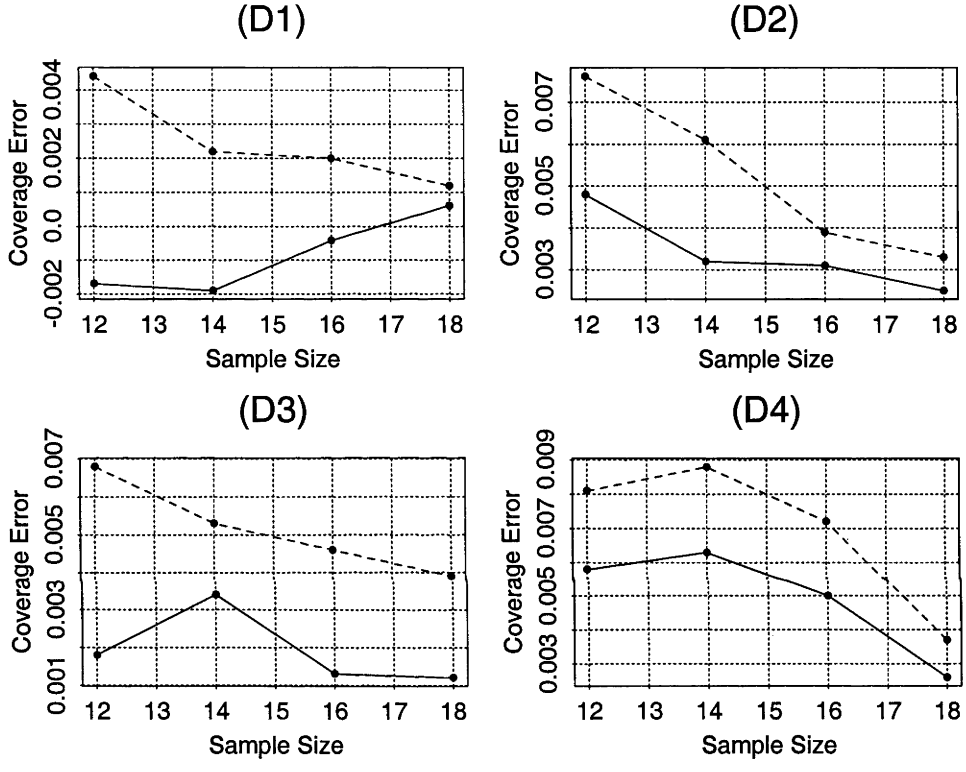

Figure 5.4: Coverage Errors. Values of coverage error for prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$ (dashed line) and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ (solid line), when the population model P is given by D1, D2, D3, and D4, the sample size n is given by 12, 14, 16, and 18, and nominal coverage $\alpha = 0.9$.

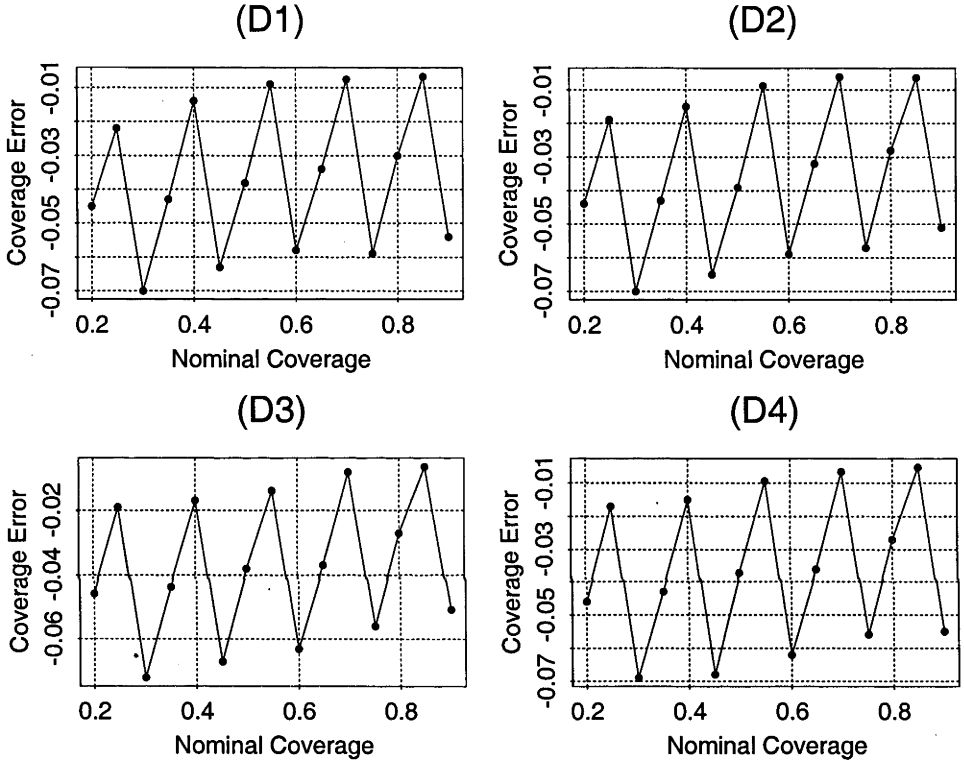

Figure 5.5: Coverage Errors. Values of coverage error for prediction interval $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$ when the population model P is given by D1, D2, D3, and D4, and the sample size $n = 12$, for various values of nominal coverage.

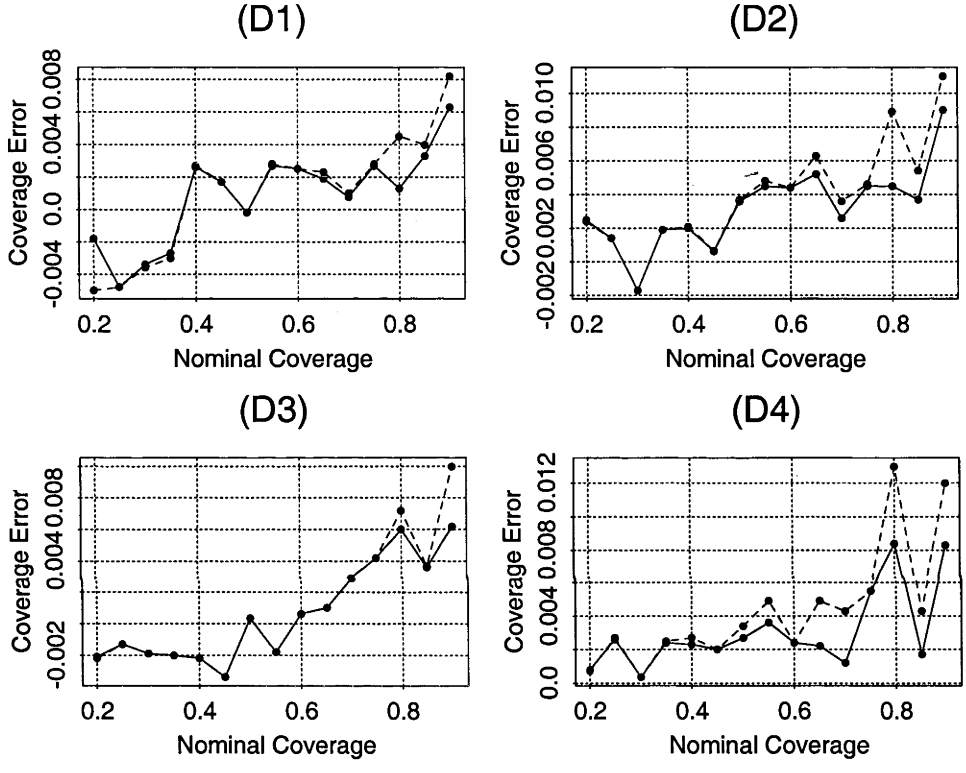

Figure 5.6: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$ (dashed line) and $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ (solid line), when the population model P is given by D1, D2, D3, and D4, and the sample size $n = 12$, for various values of nominal coverage.

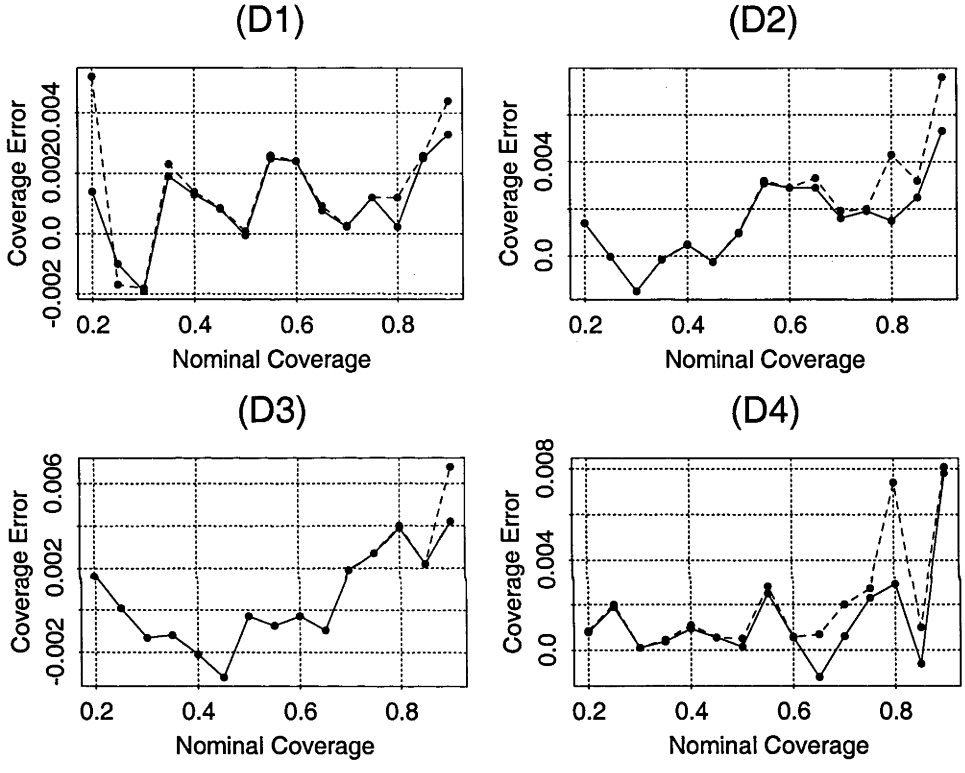

Figure 5.7: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$ (dashed line) and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ (solid line), when the population model P is given by D1, D2, D3, and D4, and the sample size $n = 12$, for various values of nominal coverage.

Appendix A

Tables

n	12	14	16	18	20	22	24	26	30	35	40	45	50
$\hat{h}_{\mathrm{N,Q2,0.90}}$	2.62	2.40	2.10	1.80	1.50	1.34	1.14	1.02	0.86	0.64	0.54	0.47	0.42
$\hat{h}_{\mathrm{N,Q2,0.95}}$	—	—	2.62	2.26	1.96	1.70	1.50	1.33	1.13	0.90	0.71	0.52	0.41
$\hat{h}_{\mathrm{N,Q3,0.90}}$	2.78	2.46	2.20	1.91	1.63	1.45	1.30	1.14	0.96	0.73	0.50	0.40	0.28
$\hat{h}_{\mathrm{N,Q3,0.95}}$	—	—	2.79	2.50	2.12	1.78	1.60	1.42	1.14	0.90	0.68	0.48	0.30

Table 1: Optimal bandwidths. Tabulated are bandwidth values that minimise absolute coverage error for a standard normal population, various values of sample size, n, and nominal coverages $\alpha = 0.9$ and 0.95. Bandwidths $\hat{h}_{\mathrm{N,Q2},\alpha}$ and $\hat{h}_{\mathrm{N,Q3},\alpha}$ are appropriate for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$ and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$, respectively.

n	12				14			
P	D1	D2	D3	D4	D1	D2	D3	D4
$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	-0.054	-0.051	-0.051	-0.055	-0.035	-0.033	-0.033	-0.033
$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	0.0082	0.011	0.010	0.012	0.0068	0.011	0.010	0.014
$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	-0.0023	0.0063	0.0022	0.0082	-0.0025	0.0054	0.0012	0.0095
$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	0.0044	0.0076	0.0068	0.0081	0.0022	0.0061	0.0053	0.0088
$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$	-0.0017	0.0048	0.0018	0.0058	-0.0019	0.0032	0.0034	0.0063
n	16				18			
P	D1	D2	D3	D4	D1	D2	D3	D4
$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	-0.018	-0.018	-0.017	-0.017	-0.0049	-0.0036	-0.0023	-0.0046
$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	0.0062	0.0090	0.0080	0.013	0.0032	0.0059	0.0059	0.0073
$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	-0.00078	0.0038	0.0020	0.0083	0.00038	0.0034	0.0021	0.0046
$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	0.0020	0.0039	0.0046	0.0072	0.0012	0.0033	0.0039	0.0037
$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$	-0.00042	0.0031	0.0013	0.0050	0.00061	0.0025	0.0012	0.0026

Table 2: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ when the population model P is given by D1, D2, D3, and D4, the sample size n is given by 12, 14, 16, and 18, and nominal coverage $\alpha = 0.9$.

D1					
α	$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$
0.20	-0.045	-0.0050	-0.0018	0.0052	0.0014
0.25	-0.022	-0.0048	-0.0048	-0.0017	-0.0010
0.30	-0.070	-0.0036	-0.0034	-0.0018	-0.0019
0.35	-0.043	-0.0030	-0.0027	0.0023	0.0019
0.40	-0.014	0.0027	0.0026	0.0014	0.0013
0.45	-0.063	0.0017	0.0017	0.00085	0.00083
0.50	-0.038	-0.00018	-0.00018	0.00010	-0.000034
0.55	-0.0090	0.0028	0.0027	0.0026	0.0025
0.60	-0.058	0.0025	0.0025	0.0024	0.0024
0.65	-0.034	0.0023	0.0019	0.00093	0.00077
0.70	-0.0075	0.0010	0.00077	0.00026	0.00024
0.75	-0.059	0.0028	0.0027	0.0012	0.0012
0.80	-0.030	0.0045	0.0013	0.0012	0.00023
0.85	-0.0067	0.0040	0.0033	0.0026	0.0025
0.90	-0.054	0.0082	0.0063	0.0044	0.0033

Table 3: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ when the population model P is given by D1 and the sample size $n = 12$, for various values of nominal coverage.

D2					
α	$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$
0.20	-0.044	0.0024	0.0025	0.0014	0.0014
0.25	-0.019	0.0014	0.0014	-0.000035	-0.000039
0.30	-0.070	-0.0017	-0.0017	-0.0015	-0.0015
0.35	-0.043	0.0019	0.0019	-0.00015	-0.00016
0.40	-0.015	0.0021	0.0020	0.00050	0.00048
0.45	-0.065	0.00066	0.00065	-0.00024	-0.00024
0.50	-0.039	0.0037	0.0036	0.0010	0.00098
0.55	-0.0089	0.0048	0.0045	0.0032	0.0031
0.60	-0.059	0.0044	0.0044	0.0029	0.0029
0.65	-0.032	0.0063	0.0052	0.0033	0.0029
0.70	-0.0062	0.0036	0.0026	0.0019	0.0016
0.75	-0.057	0.0046	0.0045	0.0020	0.0019
0.80	-0.028	0.0089	0.0045	0.0043	0.0015
0.85	-0.0065	0.0054	0.0037	0.0032	0.0025
0.90	-0.051	0.011	0.0090	0.0076	0.0053

Table 4: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ when the population model P is given by D2 and the sample size $n = 12$, for various values of nominal coverage.

D3					
α	$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$
0.20	-0.046	-0.0022	-0.0021	0.0016	0.0016
0.25	-0.019	-0.0013	-0.0013	0.000071	0.000071
0.30	-0.072	-0.0019	-0.0019	-0.0013	-0.0013
0.35	-0.044	-0.0020	-0.0020	-0.0012	-0.0012
0.40	-0.017	-0.0022	-0.0022	-0.0021	-0.0021
0.45	-0.067	-0.0034	-0.0034	-0.0032	-0.0032
0.50	-0.038	0.00034	0.00034	-0.00030	-0.00030
0.55	-0.014	-0.0018	-0.0018	-0.00075	-0.00075
0.60	-0.063	0.00061	0.00061	-0.00030	-0.00030
0.65	-0.037	0.0010	0.0010	-0.00095	-0.00095
0.70	-0.0082	0.0029	0.0029	0.0019	0.0019
0.75	-0.056	0.0042	0.0042	0.0027	0.0027
0.80	-0.027	0.0072	0.0060	0.0040	0.0039
0.85	-0.0065	0.0037	0.0036	0.0022	0.0022
0.90	-0.051	0.010	0.0062	0.0068	0.0042

Table 5: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ when the population model P is given by D3 and the sample size $n = 12$, for various values of nominal coverage.

D4					
α	$\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$	$\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$
0.20	-0.046	0.00072	0.00077	0.00084	0.00079
0.25	-0.017	0.0027	0.0026	0.0020	0.0019
0.30	-0.069	0.00035	0.00035	0.00010	0.00011
0.35	-0.043	0.0025	0.0024	0.00045	0.00037
0.40	-0.015	0.0027	0.0023	0.0011	0.00094
0.45	-0.068	0.0020	0.0020	0.00055	0.00055
0.50	-0.037	0.0034	0.0027	0.00050	0.00015
0.55	-0.0094	0.0049	0.0036	0.0028	0.0025
0.60	-0.062	0.0024	0.0024	0.00057	0.00055
0.65	-0.036	0.0049	0.0022	0.00069	-0.0012
0.70	-0.0066	0.0043	0.0012	0.0020	0.00062
0.75	-0.056	0.0055	0.0055	0.0027	0.0023
0.80	-0.027	0.013	0.0084	0.0074	0.0029
0.85	-0.0052	0.0043	0.0017	0.0010	-0.00060
0.90	-0.055	0.012	0.0083	0.0081	0.0078

Table 6: Coverage Errors. Values of coverage error for prediction intervals $\widehat{\mathcal{I}}_{\mathrm{Q1},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{AC,Q2},\alpha}$, $\widehat{\mathcal{I}}_{\mathrm{Q3},\alpha}$, and $\widehat{\mathcal{I}}_{\mathrm{AC,Q3},\alpha}$ when the population model P is given by D4 and the sample size $n = 12$, for various values of nominal coverage.

Bibliography

Aitchison, J. and I. R. Dunsmore (1975). *Statistical prediction analysis.* Cambridge: Cambridge University Press.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis.* New York: John Wiley & Sons Inc.

Bahadur, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statistics 25*, 423–462.

Bahadur, R. R. (1957). On unbiased estimates of uniformly minimum variance. *Sankhyā 18*, 211–224.

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist. 37*, 577–580.

Bai, C., P. J. Bickel, and R. A. Olshen (1990). Hyperaccuracy of bootstrap based prediction. In *Probability in Banach spaces, 7 (Oberwolfach, 1988)*, pp. 31–42. Boston, MA: Birkhäuser Boston.

Bai, C. and R. A. Olshen (1986). Discussion of paper by Peter Hall. *Ann. Statist. 16*, 953–956.

Barnard, G. A. (1949). Statistical inference. *J. Roy. Statist. Soc. Ser. B. 11*, 115–139; discussion, 139–149.

Barnard, G. A. (1985). Pivotal inference. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of Statistical Science*, Volume 6, pp. 743–746. New York: Wiley.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory.* Chichester: John Wiley & Sons Ltd. Wiley Series in Probability and Mathematical Statistics.

Barndorff-Nielsen, O. (1980). Likelihood prediction. In *Symposia Mathematica, Vol. XXV (Conf., INDAM, Rome, 1979)*, pp. 11–24. London: Academic Press.

Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika 70*(2), 343–365.

Barndorff-Nielsen, O. E. and D. R. Cox (1989). *Asymptotic techniques for use in statistics.* London: Chapman & Hall.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and asymptotics.* London: Chapman & Hall.

Barndorff-Nielsen, O. E. and D. R. Cox (1996). Prediction and asymptotics. *Bernoulli 2*(4), 319–340.

Beran, R. (1990). Calibrating prediction regions. *J. Amer. Statist. Assoc. 85*(411), 715–723.

Beran, R. (1992). Controlling conditional coverage probability in prediction. *Ann. Statist. 20*(2), 1110–1119.

Beran, R. (1993). Probability-centered prediction regions. *Ann. Statist. 21*(4), 1967–1981.

Beran, R. and P. Hall (1993). Interpolated nonparametric prediction intervals and confidence intervals. *J. Roy. Statist. Soc. Ser. B 55*(3), 643–652.

Berger, J. O. and R. L. Wolpert (1984). *The likelihood principle.* Hayward, Calif.: Institute of Mathematical Statistics.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975). *Discrete multivariate analysis: theory and practice.* The MIT Press, Cambridge, Mass.-London. With the collaboration of Richard J. Light and Frederick Mosteller.

Bjørnstad, J. F. (1990). Predictive likelihood: a review. *Statist. Sci. 5*(2), 242–265. With comments and a rejoinder by the author.

Bleistein, N. and R. A. Handelsman (1986). *Asymptotic expansions of integrals* (Second ed.). New York: Dover Publications Inc.

Bolfarine, H. and S. Zacks (1992). *Prediction theory for finite populations.* New York: Springer-Verlag.

Bowman, A., P. Hall, and T. Prvan (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika 85*(4), 799–808.

Brockwell, P. J. and R. A. Davis (1987). *Time series: theory and methods.* New York: Springer-Verlag.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probability 1*, 19–42.

Butler, R. W. (1986). Predictive likelihood inference with applications. *J. Roy. Statist. Soc. Ser. B 48*(1), 1–38. With discussion and a reply by the author.

Butler, R. W. (1989). Approximate predictive pivots and densities. *Biometrika 76*(3), 489–501.

Chow, Y. S. and H. Teicher (1997). *Probability theory* (Third ed.). New York: Springer-Verlag. Independence, interchangeability, martingales.

Chung, K. L. (1974). *A course in probability theory* (Second ed.). Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London. Probability and Mathematical Statistics, Vol. 21.

Cox, D. R. (1973). Prediction intervals and empirical Bayes confidence intervals. In J. Gani (Ed.), *Perspectives in Probability and Statistics*, London, pp. 47–55. Academic Press.

David, H. A. (1981). *Order statistics* (Second ed.). New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.

Davison, A. C. (1986). Approximate predictive likelihood. *Biometrika 73*(2), 323–332.

Davison, A. C. (1990). Correction to: "Approximate predictive likelihood" [Biometrika **73** (1986), no. 2, 323–332. *Biometrika 77*(3), 667.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press. With 1 IBM-PC floppy disk (3.5 inch; HD).

de Bruijn, N. G. (1981). *Asymptotic methods in analysis* (third ed.). New York: Dover Publications Inc.

Devroye, L. and L. Györfi (1985). *Nonparametric density estimation.* New York: John Wiley & Sons Inc. The L_1 view.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist. 7*(1), 1–26.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika 68*(3), 589–599.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* Philadelphia, Pa.: Society for Industrial and Applied Mathematics (SIAM).

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc. 82*(397), 171–200. With comments and a rejoinder by the author.

Efron, B. and G. Gong (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Statist. 37*(1), 36–48.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap.* New York: Chapman and Hall.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neerlandica 37*(2), 73–83.

Falk, M. and R.-D. Reiss (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Probab. 17*(1), 362–371.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. (London) A 144*, 285–307.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics 6*, 391–398.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist. 9*(6), 1218–1228.

Fuller, W. A. (1976). *Introduction to statistical time series.* John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.

Geisser, S. (1993). *Predictive inference.* New York: Chapman and Hall. An introduction.

Golub, G. H. and C. F. Van Loan (1996). *Matrix computations* (Third ed.). Baltimore, MD: Johns Hopkins University Press.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* New York: Springer-Verlag.

Hall, P., T. J. DiCiccio, and J. P. Romano (1989). On smoothing and the bootstrap. *Ann. Statist. 17*(2), 692–704.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application.* New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics.

Hall, P., L. Peng, and N. Tajvidi (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika 86*(4), 871–880.

Hall, P. and B. Presnell (1999). Intentionally biased bootstrap methods. *J. R. Statist. B 61*, 143–15.

Harris, I. R. (1989). Predictive fit for natural exponential families. *Biometrika 76*(4), 675–684.

Hinkley, D. (1979). Predictive likelihood. *Ann. Statist. 7*(4), 718–728.

Hinkley, D. V. (1980). Correction to: "Predictive likelihood". *Ann. Statist. 8*(3), 694.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc. 58*, 13–30.

Kabaila, P. (1993). On bootstrap predictive inference for autoregressive processes. *J. Time Ser. Anal. 14*(5), 473–484.

Kalbfleisch, J. D. and D. A. Sprott (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. Ser. B 32*, 175–208.

Kalbfleisch, J. D. and D. A. Sprott (1972). Corrigendum: "Application of likelihood methods to models involving large numbers of parameters". *J. Roy. Statist. Soc. Ser. B. 34*, 124–125.

Kreyszig, E. (1999). *Advanced engineering mathematics* (Eighth ed.). New York: John Wiley & Sons Inc.

Landers, D. and L. Rogge (1972). Minimal sufficient σ-fields and minimal sufficient statistics. Two counterexamples. *Ann. Math. Statist. 43*, 2045–2049.

Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models. *Scand. J. Statist. 1*(3), 128–134.

Lee, S. and G. A. Young (1994). Practical higher-order smoothing of the bootstrap. *Statist. Sinica 4*(2), 445–459.

Lee, S. M. S. and G. A. Young (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B Stat. Methodol. 61*(2), 353–366.

Lehmann, E. L. (1997a). *Testing statistical hypotheses* (second ed.). New York: Springer-Verlag.

Lehmann, E. L. (1997b). *Theory of point estimation.* New York: Springer-Verlag. Reprint of the 1983 original.

Magnus, J. R. and H. Neudecker (1999). *Matrix differential calculus with applications in statistics and econometrics.* Chichester: John Wiley & Sons Ltd. Revised reprint of the 1988 original.

Mathiasen, P. E. (1979). Prediction functions. *Scand. J. Statist. 6*(1), 1–21.

Miller, Jr., R. G. (1981). *Simultaneous statistical inference* (Second ed.). New York: Springer-Verlag. Springer Series in Statistics.

Mojirsheibani, M. and R. Tibshirani (1996). Some results on bootstrap prediction intervals. *Canad. J. Statist. 24*(4), 549–568.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrics 75*, 237–249.

Pitcher, T. S. (1957). Sets of measures not admitting necessity and sufficient statistics or subfields. *Ann. Math. Statist. 28*, 267–268.

Pollard, D. (1984). *Convergence of stochastic processes.* New York: Springer-Verlag.

Prakasa Rao, B. L. S. (1987). *Asymptotic theory of statistical inference.* New York: John Wiley & Sons Inc.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B. 11*, 68–84.

Rao, C. (1975). Prediction of future observations with special reference to linear models. In P. R. Krishnaiah (Ed.), *Multivariate Analysis IV*, Amsterdam, pp. 193–208. North-Holland.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Statist. 8*(2), 116–119.

Reiss, R.-D. (1989). *Approximate distributions of order statistics.* New York: Springer-Verlag. With applications to nonparametric statistics.

Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar. 4*, 191–231.

Ripley, B. D. (1987). *Stochastic simulation.* New York: John Wiley & Sons Inc.

Royden, H. L. (1988). *Real analysis* (Third ed.). New York: Macmillan Publishing Company.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.

Shao, J. and D. S. Tu (1995). *The jackknife and bootstrap.* New York: Springer-Verlag.

Shiryayev, A. N. (1984). *Probability.* New York: Springer-Verlag. Translated from the Russian by R. P. Boas.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B 43*(1), 97–99.

Silverman, B. W. (1983). Some properties of a test for multimodality based on kernel density estimates. In *Probability, statistics and analysis*, pp. 248–259. Cambridge: Cambridge Univ. Press.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman & Hall.

Silverman, B. W. and G. A. Young (1987). The bootstrap: to smooth or not to smooth? *Biometrika 74*(3), 469–479.

Simonoff, J. S. (1996). *Smoothing methods in statistics.* New York: Springer-Verlag.

Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Translation 1952*(67), 64.

Stine, R. A. (1985). Bootstrap prediction intervals for regression. *J. Amer. Statist. Assoc. 80*(392), 1026–1031.

Thombs, L. A. and W. R. Schucany (1990). Bootstrap prediction intervals for autoregression. *J. Amer. Statist. Assoc. 85*(410), 486–492.

Tierney, L., R. E. Kass, and J. B. Kadane (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc. 84*(407), 710–716.

Vidoni, P. (1995). A simple predictive density based on the p^*-formula. *Biometrika 82*(4), 855–863.

Vidoni, P. (1998). Miscellanea: A note on modified estimative prediction limits and distributions. *Biometrika 84*(4), 949–953.

Wang, S. (1989). On the bootstrap and smoothed bootstrap. *Comm. Statist. Theory Methods 18*(11), 3949–3962.

Williams, D. (1991). *Probability with martingales.* Cambridge: Cambridge University Press.

Young, G. A. (1988). Data-smoothing and bootstrap resampling. In *Model-oriented data analysis (Eisenach, 1987)*, pp. 144–150. Berlin: Springer.

Young, G. A. (1990). Alternative smoothed bootstraps. *J. Roy. Statist. Soc. Ser. B 52*(3), 477–484.